Chapter 5
Artificial Intelligence and Expert Systems

Key Topics

Turing Test
Searle's Chinese Room
Philosophical Problems in AI
Cognitive Psychology
Linguistics
Logic and AI
Robots
Cybernetics
Neural Networks
Expert Systems

5.1 Introduction

The ultimate goal of Artificial Intelligence is to create a thinking machine that is intelligent, has consciousness, has the ability to learn, has free will, and is ethical. The field involves several disciplines such as philosophy, psychology, linguistics, machine vision, cognitive science, mathematics, logic and ethics. Artificial Intelligence is a young field and the term was coined by John McCarthy and others in 1956. Alan Turing had earlier devised the Turing Test as a way to test the intelligent behaviour of a machine. There are deep philosophical problems in Artificial Intelligence, and some researchers believe that its goals are impossible or incoherent. These views are shared by Hubert Dreyfus and John Searle. Even if Artificial Intelligence is possible there are moral issues to consider such as the exploitation of artificial machines by humans, and whether it is ethical to do this. Weizembaum[1] has argued that Artificial Intelligence is unethical.

[1] Weizenbaum was a psychologist who invented the ELIZA program. This program simulated a psychologist in dialogue with a patient. He was initially an advocate of Artificial Intelligence but later became a critic.

Perhaps, one of the earliest references to creating life by artificial means is that of the classical myth of Pygmalion. Pygmalion was a sculptor who carved a woman out of ivory. The sculpture was so realistic that he fell in love with it, and offered the statue presents and prayed to Aphrodite the goddess of love. Aphrodite took pity on him and brought the statue (Galathea) to life. Pygmalion and Galathea married and had a son Paphos.

Aristotle developed syllogistic logic in the fourth century B.C. This was the first formal deductive reasoning system, and was discussed in Chapter 1. A syllogism consists of two premises (or propositions) and one conclusion. Each premise consists of two terms and there is one common middle term. The conclusion consists of the two unrelated terms from the premises. This is best illustrated by an example:

Premise 1	All Greeks are Mortal
Premise 2	Socrates is a Greek.
	....
Conclusion	Socrates is Mortal

Aristotle studied the various syllogisms and determined which were valid or invalid. Aristotle's syllogistic logic was used widely until the nineteenth century. It allowed further truths to be derived from existing truths.

Ramon Llull was a medieval writer and philosopher in the thirteenth century. He was born in Palma, Mallorca in the Balearic Islands. He did some pioneering work in computation theory by designing a method of combining attributes selected from a number of lists. He intended his method to be used as a debating tool to convert Muslims to the Christian faith through logic and reason. Llull's machine allowed a reader to enter an argument or question about the Christian faith, and the reader would then turn to the appropriate index and page to find the correct answer.

Llull's method was an early attempt to use logical means to produce knowledge. He hoped to show that Christian doctrines could be obtained artificially from a fixed set of preliminary ideas or undeniable truths. For example, the attributes of God as a supremely good, benevolent, omnipotent being were undeniable truths, and agreed between the main mono-theistic faiths of Islam, Judaism and Christianity. He believed that each field of knowledge had a limited number of basic undeniable truths, and that everything could be understood about the fields of knowledge by studying combinations of these elementary truths.

Llull's machine is called the Lullian Circle and it consists of two or more paper discs inscribed with alphabetic letters or symbols. These discs may be rotated separately to generate a large combination of ideas. A number of terms were laid around the full circumference of the circle, and repeated on an inner circle that could be rotated. These combinations were said to show all possible truths about the subject of the circle. Llull was opposed by the Grand Inquisitor of Aragon, and some of his writings were banned by the pope.

The creation of man by God is described in Genesis in the Bible. There are several stories of attempts by man to create life from inanimate objects: for example, the creation of the monster in Mary Shelly's Frankenstein. These stories go back to an earlier period: for example, there are stories of the creation of the golem in

Prague dating back to sixteenth century. The word "*golem*" is used in the Bible to refer to an incomplete substance, and the word is generally used today to describe someone who is clumsy or slow. All golems were created from mud by a very holy person who was very close to God, and who was able to create life. However, the life that the holy person could create would always be a shadow of one created by God. Golems became servants to the holy men but were unable to speak. The possession of a golem servant was a sign of the wisdom and holiness of the beholder. Other attributes of the golem were added over time, and included inscription with magical words to keep the golem animated. This also allowed the golem to be deactivated by inscribing the word "death" on the forehead of the golem.

The most famous golem story involved Rabbi Judah Loew of Prague. He was a sixteenth century rabbi and is said to have created a golem to defend the Jews of Prague from attack. According to the story the golem was created from clay on the banks of the Vltava river in Prague, and was brought to life by the Rabbi by the recitation of special incantations in Hebrew. However, as the golem grew bigger he became violent and started killing people in Prague. The Rabbi agreed to destroy the golem if the anti-semitic violence against the Jews of Prague stopped. The golem was destroyed by erasing the first letter of the word "*emet*" from the golem's forehead to make the Hebrew word "*met*", meaning death. The remains of Prague's golem are believed to be in Prague. Golems were not very bright as they were created by man and not God. Whenever, a golem was commanded to perform a task, it did so by following the instructions literally.

The golem has similarities with Mary Shelley's Frankenstein. Frankstein's monster is created by an overambitious scientist who is punished for his blasphemy of creation (in that creation is for God alone). The monster feels rejected following creation, and inflicts a horrible revenge on its creator. The story of the golum was given a more modern version in the Czech play "Rossums Universal Robots". This science fiction play by Capek appeared in Prague in 1921. It was translated into English and appeared in London in 1923. It contains the first reference to the term "robot", and the play considers the exploitation of artificial workers in a factory. The robots (or androids) are initially happy to serve humans, but become unhappy with their existence over a period of time. The fundamental question that the play is considering is whether the robots are being exploited, and if so, is this ethical, and how should the robots respond to their exploitation. It eventually leads to a revolt by the robots and the extermination of the human race. Capek considered the exploitation of non-human forms of life (the newts) in his play "War with the Newts" which was a satire on Europe and the world in the 1930s.

5.2 Descartes

Rene Descartes (Fig. 5.1) was an influential French mathematician, scientist and philosopher. He was born in a village in the Loire valley in France in 1596, and studied law at the University of Poitiers. He never practiced as a lawyer and instead served Prince Maurice of Nassau in the Netherlands. He invented the Cartesian coordinate system that is used in plane geometry and algebra. In this system, each

Fig. 5.1 Rene Descartes

point on the plane is identified through two numbers: the x-coordinate and the y-coordinate.

Descartes made important contributions to philosophy and he attempted to develop a fundamental set of principles which can be known to be true. He employed scepticism to arrive at this core set of truths, and his approach was to renounce any idea that could be doubted. He rejected the senses since the senses can deceive, and therefore they are not a sound source of knowledge. For example, during a dream the subject perceives stimuli that appear to be real, but these have no existence outside the subject's mind. Therefore, it is inappropriate to rely on one's senses as the foundation of knowledge. He argued that a powerful "evil demon or mad-scientist" could exist who sets out to manipulate and deceive subjects, thereby preventing them from knowing the true nature of reality. The evil demon could bring the subject into existence including an implanted memory. The question is how can one know for certain what is true given the limitations of the senses.

From this, Descartes deduced that there was one single principle that must be true. He argued that even if the is being deceived, then clearly he is thinking and must exist. This principle of existence or being is more famously known as "*cogito, ergo sum*" (I think, therefore I am). Descartes argued that this existence can be applied to the present only, as memory may be manipulated and therefore doubted. Further, the only existence that he sure of is that he is a "*thinking thing*". He cannot be sure of the existence of his body as his body is perceived by his senses which he has proven to be unreliable. Therefore, his mind or thinking thing is the only thing about him that cannot be doubted. His mind is used to make judgments, and to deal with un-reliable perceptions received via the senses.

Descartes constructed a system of knowledge from this one principle using the deductive method. He deduced the existence of a benevolent God using the ontological argument. This argument had first been formulated by St. Anselm in the eleventh century, and the proof of God was based on the definition and perfection of God ("The Being that nothing greater can be conceived"). Descartes argument in the fifth meditation in his Mediations on First Philosophy [Des:99] is that we have an innate idea of a supremely perfect being (God), and that God's existence may be inferred immediately from the innate idea of a supremely perfect being. The argument may be summarised as follows:

1. I have an innate idea of a supremely perfect being (i.e. God).
2. Necessarily, existence is a perfection.
3. Therefore God exists.

He then argued that as God is benevolent that he can have some trust in the reality that his senses provide, as God has provided him with a thinking mind and does not wish to deceive him. He then argued that knowledge of the external world can be obtained by both perception and deduction. In other words, he argues that reason or rationalism is the only reliable method of obtaining knowledge. Descartes arguments provided an answer to the school of scepticism that held the view that knowledge, and knowledge of existence of the external world was impossible.

Descartes was a dualist and he makes a clear mind-body distinction.[2] He states that there are two substances in the universe: mental substances and bodily substances. His proof of the existence of God and the external world are controversial and have been criticised by the sceptics and others. One well known argument against Descartes is the "the brain in the vat" argument. This is a thought experiment in which a mad scientist removes a person's brain from their body and places it in a vat and connects its neurons by wires to a supercomputer. The computer provides the disembodied brain with the electrical impulses that the brain would normally receive. The computer could then simulate reality, and the disembodied brain would have conscious experiences and would receive the same impulses as if it were inside a person's skull. There is no way to tell whether the brain is inside the vat or inside a person. The perception of a "cat" in the first case reflects the reality. However, for the second case where the brain is in the vat the perception is false, and does not correspond to reality. Since it is impossible to know whether your brain is in a vat or inside your skull it is therefore impossible to know whether your belief is valid or not.

Descartes believed that the bodies of animals are complex living machines without feelings. He dissected many animals for experiments and this included

[2] The mind-body distinction is very relevant in AI. The human mind may be considered to be a piece of software which the human brain implements. Therefore, in principle, it should be possible to code this program on a von Neumann computer yielding a mental machine. Cognitive science was originally based on this paradigm.

vivisection (i.e., the cutting up of live animals). Vivisection has become controversial in recent years on ethical grounds, and the question is whether it right that animals should be used and subject to suffering to further human interests. This debate has centred on whether animals have rights being animals given that humans have inalienable rights being human. Others have argued against vivisection claiming that it is not necessary in scientific research. Some have argued for vivisection on utilitarian grounds, with others stating that animals have rights, and should not be used as means to an end, irrespective of the expected benefits to humans.

Descartes experiments led him to believe that the actions and behaviour of non-human animals can be fully accounted for by mechanistic means, and without reference to the operations of the mind. Descartes also realised from his experiments that a lot of human behaviour is like that of animals in that it has a mechanistic explanation. For example, physiological functions and blinking are human behaviours that have a mechanistic explanation. Descartes was of the view that well-designed automata[3] could mimic many parts of human behaviour. He argued that the key differentiators between human and animal behaviour were that humans could adapt to widely varying situations, and also had the ability to use language. The use of language illustrates the power of the use of thought, and it clearly differentiates humans from animals. Animals do not possess the ability to use language for communication or reason. This, he argues, provides evidence for the presence of a soul associated with the human body. In essence, animals are pure machines, whereas humans are machines with minds (or souls).

Descartes was strongly against attributing minds to animals as it would undermine religious belief. The ascribing of thought to animals involves giving them immortal souls, since Descartes considered mental substances to be indivisible. Descartes associated the concept of mind with the Christian concept of the soul, and he argued that if animals had minds (or souls) then man's hope for an afterlife could be no more than that of flies or bats. He argued that since animals did not have minds (or souls), then this meant that they did not suffer, and that man is absolved from guilt in slaughtering and eating animals.

However, if Descartes could not infer the existence of minds in animals from their behaviour, then how could he infer the existence of minds in other humans from their behavior? His analogy could just as easily be extended to man's mind as to his body as God is capable of contriving both human and animal automata.

The significance of Descartes in the field of Artificial Intelligence is that the Cartesian dualism that humans seem to possess would need to be reflected among artificial machines. Humans seem to have a distinct sense of "I" as distinct from the body, and the "I" seems to represent some core sense or essence of being that is unchanged throughout the person's life. It somehow represents personhood, as distinct from the physical characteristics of a person that are inherited genetically. The challenge for the AI community in the longer term is to construct a machine

[3] An automaton is a self-operating machine or mechanism that behaves and responds in a mechanical way.

that (in a sense) possesses Cartesian dualism. That is, the long-term[4] goal of AI is to produce a machine that has awareness of itself as well as its environment.

5.3 The Field of Artificial Intelligence

The foundations for Artificial Intelligence goes back to the foundations of computing, and it includes the work of famous scientists, engineers and mathematicians. These include names such as Boole, Babbage, Shannon, von Neumann and Turing. There is an excellent account of the various disciplines in the AI field, as well as key problems to be solved if the field is to progress in [ONu:95]. Artificial Intelligence is a multi-disciplinary field and its many branches include:

- Computing
- Logic and Philosophy
- Psychology
- Linguistics
- Machine Vision
- Computability
- Epistemology and Knowledge representation

The British mathematician, Alan Turing, contributed to the debate concerning thinking machines, consciousness and intelligence in the early 1950s [Tur:50]. He devised the famous "Turing Test" to judge whether a machine was conscious and intelligent. Turing's paper was very influential as it raised the idea of the possibility of programming a computer to behave intelligently. Shannon considered the problem of writing a chess program in the late 1940s, and distinguished between a brute force strategy where the program could look at every combination of moves or a strategy where knowledge of chess could be used to select and examine a subset of available moves. Turing also wrote programs for a computer to play chess, and the ability of a program to play chess is a skill that is considered intelligent, even though the machine itself is not conscious that it is playing chess.

The early computer chess programs were not very sophisticated and performed poorly against humans. The modern chess programs have been quite successful, and have advantages over humans in terms of computational speed in considering combinations of moves. The chess grandmaster, Kasparov, was defeated by the IBM chess program "Deep Blue" in a six game match in 1997. The computer and Kasparov were even going into the final game which Kasparov lost. Kasparov asked for a re-match against Deep Blue, but IBM decided instead to retire the machine. Kasparov also played a 4-game match against the chess playing program X3D Fritz in 2003. The result of this match was a draw.

---

[4] This long-term goal may be hundreds of years as there is unlikely to be an early breakthrough in machine intelligence as there are deep philosophical problems to be solved. It took the human species hundreds of thousands of years to evolve to its current levels of intelligence.

The origin of the term "Artificial Intelligence" is in work done on the proposal for Dartmouth Summer Research Project on Artificial Intelligence. This proposal was written by John McCarthy (Fig. 5.2) and others in 1955, and the research project took place in the summer of 1956. Herbert A. Simon and Alan Newell did work in 1956 [NeS:56] on a program called "Logic Theorist" or "LT". This program was the world's first theorem prover, and could independently provide proofs of various theorems[5] in Russell's and Whitehead's Principia Mathematica [RuW:10]. The proof of a theorem requires creativity and intelligence, and the LT program demonstrated that computers had the ability to perform intelligent operations. Simon's and Newell's LT program was demonstrated at the Dartmouth summer conference on Artificial Intelligence. This program was very influential as it demonstrated that knowledge and information could be programmed into a computer, and that a computer had the ability to solve theorems in mathematics.

John McCarthy's approach to knowledge representation is to employ logic as the method of representing information in computer memory. He proposed a program called the Advice Taker in his influential paper "Programs with Common Sense" [Mc:59]. The idea was that this program would be able to draw conclusions from a set of premises, and McCarthy states that a program has common sense if it is capable of automatically deducing for itself "a sufficiently wide class of immediate consequences of anything it is told and what it already knows".

The Advice Taker aims to use logic to represent knowledge (i.e., premises that are taken to be true), and it then applies the deductive method to deduce further truths from the relevant premises.[6] That is, the program manipulates the formal language (e.g., predicate logic), and provides a conclusion that may be a statement or an imperative. McCarthy's Advice Taker differs from Simon's and Newell's LT program in that the latter program is a heuristic for deduction. McCarthy envisaged that the Advice Taker would be a program that would be able to learn and improve. This would be done by making statements to the program, and telling it about its symbolic environment. The program will have available to it all the logical consequences of what it has already been told and previous knowledge. McCarthy's desire was to create programs to learn from their experience as effectively as humans do.

McCarthy's philosophy is that common sense knowledge and reasoning can be formalised with logic. A particular system is described by a set of sentences in logic. These logic sentences represent all that is known about the world in general, and what is known about the particular situation and the goals of the systems. The

[5] Russell is reputed to have remarked that he was delighted to see that the Principia Mathematica could be done by machine, and that if he and Whitehead had known this in advance that they would not have wasted 10 years doing this work by hand in the early twentieth century. The LT program succeeded in proving 38 of the 52 theorems in Chapter 2 of Principia Mathematica. Its approach was to start with the theorem to be proved and to then search for relevant axioms and operators to prove the theorem.

[6] Of course, the machine would somehow need to know what premises are relevant and should be selected for to apply the deductive method from the many premises that are already encoded.

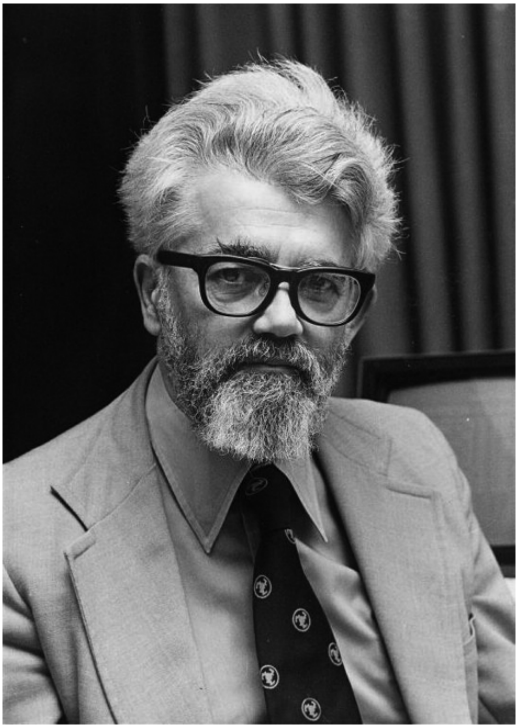

Fig. 5.2 John McCarthy
Courtesy of John McCarthy.

program then performs actions that it infers are appropriate for achieving its goals. That is, common sense[7] knowledge is formalised by logic, and common sense problems are solved by logical reasoning.

5.3.1 Turing Test and Strong AI

Turing contributed to the debate concerning artificial intelligence in his 1950 paper on Computing, machinery and intelligence [Tur:50]. Turing's paper considered whether it could be possible for a machine to be conscious and to think. He predicted that it would be possible to speak of machines thinking and he devised a famous experiment that would allow a computer to be judged as a conscious and thinking machine. This is known as the "Turing Test". The test itself is an adaptation of a well-known party game which involves three participants. One of them, the judge, is placed in a separate room from the other two: one is a male and the other is a female. Questions and responses are typed and passed under the door. The objective

[7] Common sense includes basic facts about events, beliefs, actions, knowledge and desires. It also includes basic facts about objects and their properties.

of the game is for the judge to determine which participant is male and which is female. The male is allowed to deceive the judge whereas the female is supposed to assist.

Turing adapted this game by allowing the role of the male to be played by a computer. If the judge could not tell which of the two participants was human or machine, then the computer could be considered intelligent. The test involves a judge who is engaged in a natural language conversation with two other parties, one party is a human and the other is a machine. If the judge cannot determine which is machine and which is human, then the machine is said to have passed the "Turing Test". That is, a machine that passes the Turing Test must be considered intelligent, as it is indistinguishable from a human. The test is applied to test the linguistic capability of the machine rather than the audio capability, and the conversation is limited to a text only channel.

Turing's work on "thinking machines" caused a great deal of public controversy as defenders of traditional values attacked the idea of machine intelligence. Turing's paper led to a debate concerning the nature of intelligence. There has been no machine developed, to date, that has passed the Turing test.

Turing strongly believed that machines would eventually be developed that would stand a good chance of passing the "Turing Test". He considered the operation of "thought" to be equivalent to the operation of a discrete state machine. Such a machine may be simulated by a program that runs on a single, universal machine, i.e. a computer.

Turing viewpoint that a machine will one day pass the Turing Test and be considered intelligent is known as "*strong artificial intelligence*". It states that a computer with the right program would have the mental properties of humans. There are a number of objections to strong AI, and one well-known rebuttal is that of Searle's Chinese Room argument [Sea:80].

The Chinese Room argument is a thought experiment that is used to demonstrate that a machine will never have the same cognitive qualities as a human. A man is placed into a room into which Chinese writing symbols are given to him. He has the knowledge of what symbol to use to respond to each symbol that is presented to him. However, he has no idea as to what each symbol means. Essentially, he is communicating with the person who is giving the symbols to him, and answering any questions that are posed, without the slightest understanding of what he is doing and what the symbols mean.

The closed room has two slots where slot 1 is used to input the Chinese characters. The person in the room has no understanding of what these Chinese characters mean. However, a rulebook is provided that allows new Chinese character to be created from the Chinese characters which have already been input. Slot 2 is used to output the Chinese characters. There are essentially three steps in the process:

1. Chinese characters are entered through slot 1.
2. The rulebook is employed to construct new Chinese characters.
3. Chinese characters are outputted to slot 2.

The process is essentially that of a computer program which has an input; performs a computation based on the input; and then finally produces an output. Further, the rulebook is such that people outside the room are able to send questions such as "How are you?". The responses to these questions are provided following the rulebook, and meaningful answers are provided. That is, the computer program simulates a human being who understands Chinese, even though the person has not the slightest understanding of the language.

The question "Do you understand Chinese?" could potentially be asked, and the rulebook would be consulted to produce the answer "Yes, of course" that is, despite of the fact that the person inside the room has not the faintest idea of what is going on, it will appear to the person outside the room that the person inside is knowledgeable on Chinese. The person in the room is just following rules without understanding.

Searle has essentially constructed a machine which can never be mental. Changing the program essentially means changing the rulebook, and this does not increase understanding. The strong artificial intelligence thesis states that given the right program, *any* machine running it would be mental. However, Searle argues that the program for this Chinese room would not understand anything, and that therefore the strong AI thesis must be false. In other words, Searle's Chinese room argument is a rebuttal of strong AI by showing that a program running on a machine that appears to be intelligent has no understanding whatsoever of the symbols that it is manipulating. That is, given any rulebook (i.e., program), the person would never understand the meanings of those characters that are manipulated.

Searle argument essentially states that just because the machine acts like it knows what is going on, it only knows what it is programmed to know. It differs from humans in that it is not aware of the situation like humans are. The question is whether a machine can only do what it is told to do, or whether it may think for itself and have consciousness. Searle's argument suggests that machines may not have intelligence or consciousness, and the Chinese room argument applies to any Turing equivalent computer simulation.

There are several rebuttals of Searle's position[8] and one well-known rebuttal attempt is the "System Reply" argument. This reply argues that if a result associated with intelligence is produced, then intelligence must be found somewhere in the system. The proponents of this argument draw an analogy between the human brain and its constituents. None of its constituents have intelligence but the system as a whole (i.e., the brain) exhibits intelligence. Similarly, the parts of the Chinese room may lack intelligence, but the system as a whole is intelligence. However, this rebuttal argument has been criticized by others as begging the question.

[8] I don't believe that Searle's argument proves that Strong AI is impossible. However, I am not expecting to see intelligent machines anytime soon.

5.4 Philosophy and AI

Artificial Intelligence includes the study of knowledge and the mind, and there are deep philosophical problems to be solved in the field. The philosophy of epistemology and the philosophy of mind are concerned with fundamental questions such as:

- What is being?
- What is knowledge?
- What is mind?
- What is consciousness?

Early work on philosophy was done by the Greeks as they attempted to understand the world and the nature of being and reality.[9] Thales and the Miletians[10] attempted to find an underlying principle: e.g., water, or for other Miletians, earth, air, wind and fire, where such a principle that would explain everything. Pythagoras believed that mathematical numbers were the basic underlying principle, and that everything (e.g., music) could be explained in terms of number. Plato distinguished between the world of appearances and the world of reality. He argued that the world of appearances resembles the flickering shadows on a cave wall, whereas reality is in the world of ideas[11] or forms, in which objects of this world somehow participate. Aristotle wondered how many forms there are: for example, is there a separate form for a dog, breed of dog, etc. Aristotle proposed the framework of a substance which includes form plus matter. For example, the matter of a wooden chair is the wood that it is composed of, and its form is the general form of a chair. The problem of being is relevant to the mind-body problem in AI.

Acquinus was a medieval scholastic philosopher who was deeply influenced by Aristotle.[12] Acquinus distinguished between matter and form and also considered questions such as existence and understanding. Modern Thomists,[13] such as Bernart Lonergan have developed his ideas into the nature of understanding and cognition. Lonergans's study of cognition [Lon:58] was influenced by the eureka step in the discovery of Archimedes, and also the in the sudden insight in which the solution of a mathematical puzzle comes about. Lonergan outlines the steps associated with the sudden insight that comes in solving a mathematical puzzle or scientific discovery. Artificial Intelligence requires a theory of cognition if it is to make serious progress.

---

[9] The study of philosophical questions related to being is referred to as Ontology (or Metaphysics). The study of philosophical questions related to knowledge is referred to as epistemology.

[10] The term "Miletians" refers to inhabitants of the Greek city state Miletus which is located in modern Turkey. Anaximander and Anaximenes were two other Miletians who made contributions to early Greek philosophy approx 600 BC.

[11] Plato was an Idealist: i.e., that reality is in the world of ideas rather than the external world. Realists (in contrast) believe that the external world corresponds to our mental ideas.

[12] He refers to Aristotle as "The Philosopher" in Sumna Theologicae.

[13] The term "Thomist" denotes a follower of the philosophy of St. Thomas Acquinus.

Descartes was discussed earlier and his influence on the philosophy of mind and AI is significant. He was deeply concerned with a firm foundation for knowledge and being, and his approach was to doubt everything that could be doubted except his own existence. He deduced his own existence from the fact that he could think: i.e., *Cogito, ergo sum.*[14] This thinking thing (*res cogitans* or mind/soul) was distinct from the rest of nature and interacted with the world through the senses to gain knowledge. Knowledge was gained by mental operations using the deductive method, where starting from the premises that are known to be true, further truths could be logically deduced. Descartes founded what would become known as the Rationalist school of philosophy where knowledge was derived solely by human reasoning. He distinguished between the mind and the body (Cartesian dualism), and the analogy of the mind in AI would be the central processor unit (or an AI program) of a computer, with knowledge gained by sense perception by the computer hardware (e.g., machine vision) and logical deduction.

British Empiricism rejected the Rationalist position, and stressed the importance of empirical data in gaining knowledge about the world. Its philosophy argued that all knowledge is derived from sense experience. It consisted of philosophers such as Hobbes, Locke, Berkeley[15] and Hume. Locke argued that a child's mind is a blank slate (*tabula rasa*) at birth, and that all knowledge is gained by sense experience. Berkeley (Fig. 5.3) considered the frame problem: i.e., the fact that as we move around a particular object its perspective changes, and yet we as humans still manage to recognise it as the same object. This problem is relevant in machine vision in AI.

Fig. 5.3 George Berkely Bishop of Cloyne

[14] I think, therefore I am.

[15] Berkeley was an Irish philosopher (not British). He was born in Dysart castle in Kilkenny, Ireland; educated at Trinity College, Dublin; and served as bishop of Cloyne in Co. Cork. He planned to establish a seminary in Bermuda for the sons of colonists in America, but the project failed due to lack of funding from the British government. Berkeley University in San Francisco is named after him.

Berkeley appealed to ideas like the soul to unify the various appearances of the chair as the observer moves around the room, and to God to keep in existence things that were not being perceived.[16] Berkely argued that the ideas in a man's mind have no existence outside his mind [Ber:99], and this philosophical position is known as Idealism. David Hume formulated the standard empiricist philosophical position in his book "An Enquiry concerning Human Understanding" [Hum:06], and this was published in 1748.

Hume (Fig. 5.4) distinguished between ideas and impressions. The term "*impression*" means the lively perceptions that a human directly experiences when hearing, seeing, experiencing anger, and so on. Ideas are less lively perceptions and are experienced when reflecting on sensations such as hearing, seeing and so on. Hume argued that every idea is copied from some impression, and that it is not possible for an idea to arise without its corresponding impression. Hume identified three relations between ideas: resemblance (i.e., similarity), contiguity (experienced together or otherwise) in time or space, and cause or effect.

The empiricist position formulated by Hume argued that all objects of human knowledge may be divided into two kinds (Table 5.1):

Table 5.1 Humes theory of empiricism

Kind of knowledge	Description
Relations of Ideas	These include the truths of propositions in Geometry, Algebra and Arithmetic. They are demonstrated by abstract reasoning through the operation of the human mind. These truths are independent of the outside world, and their denial leads to a contradiction.
Matter of Fact	The evidence for the truth of this kind of knowledge is not as conclusive as knowledge gained through relations of ideas. These truths may not be demonstrated (logically) to be true, and the denial of these propositions does not lead to a contradiction. The evidence for these truths is generally based on experience, and cause and effect. There is a chain of reasons for believing that a particular proposition is true, and these reasons are based on experience.

Hume argued that all knowledge consists of either matters of fact or propositions that may be demonstrated (such as geometry) via deduction reasoning in the operations of the mind.

[16] Berkeley's theory of Ontology is that for an entity to exist it must be perceived: i.e., "*Esse est percipi*". He argues that "It is an opinion strangely prevailing amongst men, that houses, mountains, rivers, and in a world all sensible objects have an existence natural or real, distinct from being perceived".
This led to a famous Limerick that poked fun at Berkeley's theory. "There once was a man who said God; Must think it exceedingly odd; To find that this tree, continues to be; When there is no one around in the Quad".
The reply to this Limerick was appropriately: "Dear sir, your astonishments odd; I am always around in the Quad; And that's why this tree will continue to be; Since observed by, yours faithfully, God".

Fig. 5.4 David Hume

The denial or a matter of fact does not involve a contradiction, and matter of fact propositions may be demonstrated by arguments based on experience or cause and effect. They are based entirely based on experience, and experience provides insight into the nature and bounds of cause and effect, and enables the existence of one object to be inferred from that of another.

Hume [Hum:06] argued that any subject[17] proclaiming knowledge that does not adhere to these empiricist principles should be committed to the flames[18] as such knowledge contains nothing but sophistry and illusion.

Kant's Critique of Pure Reason [Kan:03] was published in 1781 and is a response to Hume's theory of empiricism. Kant argued that there is a third force in human knowledge that provides concepts that can't be inferred from experience. Such concepts include the laws of logic (e.g., modus ponens), causality, and so on, and Kant argued that the third force was the manner in which the human mind structures its experiences. These structures (or systematic rules by which the mind structures its experience) are called categories, and include entities such as modus ponens.

[17] Hume argues that these principles apply to subjects such as Theology as its foundations are in faith and divine revelation which are neither matters of fact nor relations of ideas.

[18] "When we run over libraries, persuaded of these principles, what havoc must we make? If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: for it can contain nothing but sophistry and illusion".

The continental school of philosophy included thinkers such as Heidegger, and Merleau-Ponty who argued that the world and the human body are mutually intertwined. Heidegger emphasized that existence can only be considered with respect to a changing world. Merleau-Ponty emphasized embodiment: i.e., that human cognition relies heavily on the body, and he identified the concept of a body-subject that actively participates both as the perceiver of knowledge and as an object of perception.

The analytic school of philosophy was an attempt in the early twentieth century to clean up philosophy. The school believed that many of the philosophical problems were due to inadequate scientific knowledge and an abuse of language. The school was founded in Vienna after the first-world war, and it was known as the Vienna circle. They developed what became known as logical empiricism and this employed the verification principle. This principle stated that a proposition could have meaning if and only if it could be verified. This principle was later shown to be flawed and logical empiricism failed.

Wittgenstein was living in Vienna at the time, and although he was not a member of the circle, he had done important work in his Tractatus Logico Philosophicus [Wit:22] to produce a full deconstruction of language into atomic propositions. These atomic propositions refer to simple objects in the world, which could then be combined to form more complex objects.

Philosophy has been studied for over two millennia but to date very little progress has been made in solving the fundamental philosophical questions. Artificial Intelligence will need to take a practical approach to dealing with the key philosophical questions.

5.5 Cognitive Psychology

Psychology arose out of the field of psychophysics in the late nineteenth century in which various German pioneers attempted to quantify perception and sensation. In fact, one of these pioneers, Fechner, produced a mathematical formulation of the relationship between stimulus and sensation:

$$S = k \log I + c$$

The symbol S refers to the intensity of the sensation, the symbols k and c are constants, and the symbol I refers to the physical intensity of the stimulus. Psychology was defined by William James as the science of mental life.

One of the early behaviouralist psychologist's was Pavlov who showed that it was possible to develop a conditional reflex in a dog. His experiment showed that it was possible to make a dog salivate in response to the ringing of a bell. He did this by ringing a bell each time before meat was provided to the dog, and the dog therefore associated the presentation of meat with the ringing of the bell after a training period.

Skinner developed the concept of conditioning further using rewards to reinforce desired behaviour, and punishment to discourage undesired behaviour. The reinforcement of desired behaviour helps to motivate the individual to behave in a desired way, with punishment used to deter the individual from performing undesired behaviour. The behavioural theory of psychology explains many behavioural aspects of the world, although it does not really explain aspects of complex learning such as language development. Further, behaviouralism does not take human experience into account, and experience plays an important role in explaining human behaviour.

Merleau-Ponty[19] considered the problem of what the structure of the human mind must be in order to allow the objects of the external world to exist in our minds in the form that they do. He argued for phenomenology as originally developed by Hegel and Husserl. This involves a focus and exploration of phenomena with the goal of establishing the essential features of experience. Merleau-Ponty developed the concept of phenomenology further by introducing the concept of the body-subject. This concept is distinct from the Cartesian view that the world is just an extension of our own mind, and instead takes the position that the world and the human body are mutually intertwined. That is, in the Cartesian viewpoint it is first necessary for the self to be aware of and to recognize its own existence prior to being aware of and recognizing the existence of anything else. The body-subject concept expresses the embodied feeling of reality that all humans share.

Merleau-Ponty argued that the body plays a key role in the process of human understanding. The body has the ability to perceive the world and it plays a double role in that it is both the subject (i.e., the perceiver) and the object (i.e., the entity being perceived) of experience. Human understanding and perception is dependent on the body's capacity to perceive via the senses, and its ability to interrogate its environment. Merleau-Ponty argued that there is a symbiotic relationship between the perceiver and what is being perceived, and he argues that as our consciousness develops the self imposes richer and richer meanings on objects. He provides a detailed analysis of the flow of information between the body-subject and the world.

Cognitive psychology is a branch of psychology that is concerned with learning, language, memory and internal mental processes. Its roots lie in the child development psychology developed by Piaget, and in the Gesalt psychology developed by Wertheimer. The latter argues that the operations of the mind are holistic, and that the mind contains a self-organising mechanism. The philosophical position known as Holism argues that the sum of the parts is less than the whole subject, and it is the opposite of the philosophical position of logical Atomism[20] developed by Bertrand Russell. Russell (and also Wittgenstein) attempted to identify the atoms of thought: i.e., the elements of thought that cannot be divided into smaller pieces. The philosophical position of logical atomism was that all truths are ultimately dependent on

[19] Merleau-Ponty was a French philosopher who was strongly influenced by the phenomenology of Husserl. He was also closely associated with the French existentialist philosophers such as Jean-Paul Sartre and Simone De Beauvoir.

[20] Atomism actually goes back to the work of the ancient Greeks and was originally developed by Democritus and his teacher Leucippus in the fifth century BC. Atomism was rejected by Plato in the dialogue the Timaeus.

a layer of atomic facts. It had an associated methodology whereby by a process of analysis it attempted to construct more complex notions in terms of simpler ones.

Cognitive psychology is concerned with how people understand, diagnose, and solve problems, and the mental processes that take place during a stimulus and its corresponding response. It argues that solutions to problems take the form of rules, heuristics and sudden insight. It was developed in the late 1950s, and it views the mind as having a certain conceptual structure. Cognition is considered as the processes by which the sensory input is transformed and used. The dominant paradigm in the field has been the information processing model, and this viewpoint considers the mental processes of thinking and reasoning as being equivalent to the software running on the computer: i.e., the brain. The information processing model has associated theories of input, representation of knowledge, processing of knowledge and output.

Cognitive psychology has been applied to artificial intelligence from the 1960s, and it has played an important role in the integration of various approaches to the mind and mental processes. Some of the research areas in cognitive psychology include:

- Perception
- Concept Formation
- Memory
- Knowledge Representation
- Learning
- Language
- Grammar and Linguistics
- Thinking
- Logic and Problem Solving

The success of these research areas in cognitive psychology will strongly influence developments in Artificial Intelligence. For a machine to behave with intelligence it must be able to perceive objects of the physical world and form concepts about the world. It must be able to remember knowledge that it has already been provided with, as well as possessing an understanding of temporal events. There must be a way to represent knowledge in a machine that allows the knowledge to be easily retrieved and used in analysis and decision making. For a machine to exhibit intelligence it must possess an understanding of language. That is, it must be capable of understanding written language as well as possessing the ability to communicate in written or audio form. A thinking machine must be capable of thought, analysis and problem solving.

5.6 Linguistics

Linguistics is the theoretical or applied scientific study of language. Theoretical linguistics is concerned with the study of syntax and semantics, whereas applied linguistics is concerned with putting linguistic theories into practice.

The study of syntax includes the study of the rules of grammar, and the grammar determines the way by which valid sentences and phrases are produced in the language. It also includes the study of morphology and phonetics. Morphology is concerned with the formation and alteration of words, and phonetics is concerned with the study of sounds, and how sounds are produced and perceived as speech (or non-speech).

Computational Linguistics is the study of the design and analysis of natural language processing systems. It is an interdisciplinary field and includes linguists, computer scientists, experts in artificial intelligence, cognitive psychologists, and mathematicians. Human language is highly complex, and this is reflected in the sophistication of human intelligence.

Computational linguistics originated in the United States in the 1950s in an effort to design software to translate texts written in foreign languages into English. The objective at that time was to develop an automated mechanism by which Russian language texts could be translated directly into English without human intervention. It was believed that given the speed and power of the emerging computer technology that it was only a matter of time before it would be possible by design software to analyse Russian texts and to automatically provide a corresponding English translation.

However, the initial results of automated machine translation of human languages were not very successful, and it was realised that the automated processing of human languages was considerably more complex than initially believed. This led to the birth of a new field called computational linguistics, and the objective of this field is to investigate and to develop algorithms and software for processing natural languages. It is a sub-field of artificial intelligence, and deals with the comprehension and production of natural languages.

The task of translating one language into another requires an understanding of the grammar of both languages. This includes an understanding of the syntax, the morphology, semantics and pragmatics of the language. For Artificial Intelligence to become a reality it will need to make major breakthroughs in computational linguistics.

5.7 Cybernetics

The interdisciplinary field of cybernetics[21] began in the late 1940s when concepts such as information, feedback, and regulation were generalized from engineering to other systems. These include systems such as living organisms, machines, robots and language. The term "*cybernetics*" was coined by Norma Wiener, and it taken from the Greek word "κυβερνήτης" (meaning steersman or governor).

[21] Cybernetics was defined by Couffignal (one of its pioneers) as the art of ensuring the efficacy of action.

The name is appropriate as a steersman has a set of goals in traveling to a particular destination, and will need to respond to different conditions on the journey. Similarly, the field of cybernetics is concerned with an interaction of goals, predictions, actions, feedback, and response in systems of all kinds. It is the study of communications and control and feedback in living organisms and machines to ensure efficient action. Practitioners of cybernetics use models of organisations, feedback and goals to understand the capacity and limits of any system.

Cybernetics is concerned with approaches to acquire knowledge through control and feedback, and it is a branch of applied epistemology. The principles of cybernetics are similar to the way that humans acquire knowledge (e.g., learning is achieved through a continuous process of transformation of behaviour through feedback from parents and peers, rather than an explicit encoding of knowledge). The systems studied include observed systems and observing systems (systems that explicitly incorporate the observer into the system).

AI is based on the assumption that knowledge is something that can be stored inside a machine, and that the application of stored knowledge to the real world in this way constitutes intelligence. External objects are mapped to internal states on the machine, and machine intelligence is manifested by the manipulation of the internal states. This approach has been reasonably successful with rule-based expert systems. However, the progress of AI in creating intelligent machines has not been very successful to date, and an alternative approach based on cybernetics may become more important in the future.

Cybernetics views information (or intelligence) as an attribute of an interaction, rather than something that is stored in a computer. The concepts of cybernetics have become part of other disciplines, and it may have an important role to play in the Artificial Intelligence field.

5.8 Logic and AI

Logic is a key discipline in the AI field, and the goal is to use mathematical logic to formalise knowledge and reasoning to enable AI problems to be solved. John McCarthy and others have argued that mathematical logic has a key role to play in the formalisation of common-sense knowledge and reasoning. Common-sense knowledge includes basic facts about actions and their effects, facts about beliefs and desires, and facts about knowledge and how it is obtained. He argues that common-sense reasoning is required for solving problems in the real world, and that therefore it is reasonable to apply common-sense knowledge to the world of artificial intelligence. His approach is to use mathematical logic to formalise common-sense knowledge to allow common-sense problems to be solved by logical reasoning.

This requires sufficient understanding of the common-sense world to enable facts to be formalised. One problem is that often the reasoner doesn't know which facts are relevant in solving a particular problem. Knowledge that was thought to be irrelevant to the problem to be solved may be essential. A computer may have millions of facts stored in its memory, and the problem is how can it determine the relevant facts from its memory to serve as premises for logical deduction.

McCarthy philosophy of producing programs with common sense was discussed in his influential 1959 paper [Mc:59]. This paper discusses various common-sense problems such as getting home from the airport. Other examples of common-sense problems are diagnosis, spatial reasoning, and understanding narratives that include temporal events.

Computers need a precise language to store facts about the world and to reason about them. It needs a rigorous definition of the reasoning that is valid as this will determine how the deduction of a new formula from a set of formulae is done. Mathematical logic is the standard approach to express premises, and it includes rules of inferences that are used to deduce valid conclusions from a set of premises.

The propositional and predicate calculii were discussed earlier in Chapter 4. Propositional calculus associates a truth-value with each proposition, and includes logical connectives to produce formulae such as $A \Rightarrow B$, $A \wedge B$, and $A \vee B$. The truth values of the propositions are normally the binary values of *true* and *false*. There are other logics, such as 3-valued logic or fuzzy logics that allow more than two truth-values for a proposition. Predicate logic is more expressive than propositional logic, and it can formalise the syllogism "All Greeks are mortal; Socrates is a Greek; Therefore, Socrates is mortal". The predicate calculus consists of:

- Axioms
- Rules for defining well-formed formulae
- Inference rules for deriving theorems from premises

A formula in predicate calculus is built up from the basic symbols of the language. These include variables, predicate symbols such as equality; function symbols, constants logical symbols such as: $\exists$, $\wedge$, $\vee$, $\neg$, and punctuation symbols such as brackets and commas. The formulae of predicate calculus are built from terms, where a *term* is defined recursively as a variable or individual constant or as some function containing terms as arguments. A formula may be an atomic formula, or built from other formulae via the logical symbols.

There are several rules of inference associated with predicate calculus, and the most important of these are modus ponens and generalisation. The rule of modus ponens states that given two formulae p, and $p \Rightarrow q$, then we may deduce q. The rule of generalisation states that given a formula p that we may deduce $\forall(x)p$.

McCarthy's approach to programs with common sense has been criticized by Bar-Hillel and others on several grounds:

- How can a computer know from the set of facts that it knows which facts are appropriate to apply in deduction in a given situation.
- Many philosophers believe that common sense is fairly elusive.

However, while logic is unlikely to prove the main solution to the problems in AI, it is likely to form part of the solution.

5.9 Computability, Incompleteness and Decidability

An algorithm (or procedure) is a finite set of unambiguous instructions to perform a specific task. The term "algorithm" is named after the Persian mathematician Al-Khwarizmi. The concept of an algorithm was defined formally by Church in 1936 with the lambda calculus, and independently by Turing with the Turing machines. Turing defined computability in terms of the mechanical procedure of a Turing machine, whereas Church defined computability in terms of the lambda calculus. Later it was shown that computability by Turing machines and lambda calculus are equivalent.

Formalism was proposed by Hilbert as a foundation for mathematics in the early twentieth century. A formal system consists of a formal language, a set of axioms and rules of inference. Hilbert's programme was concerned with the formalization of mathematics (i.e., the axiomatization of mathematics) together with a proof that the axiomatization was consistent. The specific objectives of Hilbert's programme were to:

- Develop a formal system where the truth or falisty of any mathematical statement may be determined.
- A proof that the system is consistent (i.e., that no contradictions may be derived).

A proof in a formal system consists of a sequence of formulate, where each formula is either an axiom or derived from one or more preceding formulae in the sequence by one of the rules of inference. Hilbert believed that every mathematical problem could be solved and therefore expected that the formal system of mathematics would be completes (that is, all truths could be proved within the system) and decidable: i.e., that the truth or falisty of any mathematical proposition could be determined by an algorithm.

Russell and Whitehead published Principia Mathematica in 1910, and this three-volume work on the foundations of mathematics attempted to derive all mathematical truths in arithmetic from a well-defined set of axioms and rules of inference. The questions remained whether the Principia was *complete* and *consistent.* That is, is it possible to derive all the truths of arithmetic in the system and is it possible to derive a contradiction from the Principia's axioms?

Gödel's second incompleteness theorem [Goe:31] showed that first order arithmetic is incomplete, and that the consistency of first-order arithmetic cannot be proved within the system. Therefore, if first-order arithmetic cannot prove its own consistency, then it cannot prove the consistency of any system that contains first-order arithmetic.

Hilbert also believed that formalism would be decidable: there would be a mechanical procedure (or algorithm) to determine whether a particular statement was true or false. However, Church and Turing independently showed this to be impossible in 1936. The only way to determine whether a statement is true or false is to try to solve it.

5.10 Robots

The first use of the term "robot" was by the Czech playwright Karel Capek in his play "*Rossum's Universal Robots*" performed in Prague in 1921. The word "robot" is from the Czech word for forced labour. The play was discussed earlier in Section 5.5, and the theme explored is whether it is ethical to exploit artificial workers in a factory, and what response the robots should make to their exploitation. Capek's robots were not mechanical or metal in nature and were instead created through chemical means. Capek, in fact, rejected the idea that machines created from metal could think of feel.

The science fiction writer Asimov wrote several stories about robots in the 1940s including the story of a robotherapist.[22] He predicted the rise of a major robot industry, and he also introduced a set of rules (or laws) which stipulated the good behaviors that robots are expected to observe. These are known as the three Laws of Robotics and a fourth law was later added by Asimov (Table 5.2).

The term "robot" is defined by the Robot Institute of America as:

Definition 5.1 (Robots) A re-programmable, multifunctional manipulator designed to move material, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks.

Joseph Engelberger and George Devol are considered the fathers of robotics. Engelberger set up the first manufacturing company "Unimation" to make robots, and Devol wrote the necessary patents. Their first robot was called the "Unimate". These robots were very successful and reliable, and saved their customer (General Motors) money by replacing staff with machines. The robot industry continues to play a major role in the automobile sector.

Robots have been very effective at doing clearly defined repetitive tasks, and there are many sophisticated robots in the workplace today. These robots are industrial manipulators and are essentially computer controlled "arms and hands". However, fully functioning androids are many years away.

Table 5.2 Laws of Robotics

Law	Description
Law Zero	A robot may not injure humanity, or, through inaction, allow humanity to come to harm.
Law One	A robot may not injure a human being, or, through inaction, allow a human being to come to harm, unless this would violate a higher order law.
Law Two	A robot must obey orders given it by human beings, except where such orders would conflict with a higher order law.
Law Three	A robot must protect its own existence as long as such protection does not conflict with a higher order law.

[22] The first AI therapist was the ELIZA program produced by Weizenbaum in the mid-1960s.

Robots can also improve the quality of life for workers as they can free human workers from performing dangerous or repetitive jobs. Further, it leads to work for robot technicians and engineers. Robots provide consistently high-quality products and can work tirelessly 24 hours a day. This helps to reduce the costs of manufactured goods thereby benefiting consumers. They do not require annual leave but will, of course, from time to time require servicing by engineers or technicians. However, there are impacts on workers whose jobs are displaced by robots.

5.11 Neural Networks

The term "neural network" (artificial or biological) refers to an interconnected group of processing elements called nodes or neurons. These neurons cooperate and work together to produce an output function. Neural networks may be artificial or biological. A biological network is part of the human brain, whereas an artificial neural network is designed to mimic some properties of a biological neural network. The processing of information by a neural network is done in parallel rather than in series. A unique property of a neural network is fault tolerance: i.e., it can still perform (within certain tolerance levels) its overall function even if some of its neurons are not functioning. There are trainable neural network systems that can learn to solve complex problems from a set of examples. These systems may also use the acquired knowledge to generalise and solve unforeseen problems.

A biological neural network is composed of billions of neurons (or nerve cells). A single neuron may be physically connected to thousands of other neurons, and the total number of neurons and connections in a network may be extremely large. The human brain consists of many billions of neurons, and these are organized into a complex intercommunicating network. The connections are formed through axons[23] to dendrites,[24] and the neurons can pass electrical signals to each other. These connections are not just the binary digital signals of *on* or *off*, and instead the connections have varying strength which allows the influence of a given neuron on one of its neighbors to vary from very weak to very strong.

That is, each connection has an individual weight (or number) associated with it that indicates its strength. Each neuron sends its output value to all other neurons to which it has an outgoing connection. The output of one neuron can influence the activations of other neurons causing them to fire. The neuron receiving the connections calculates its activation by taking a weighted sum of the input signals. The output is determined by the activation function based on this activation. Networks learn by changing the weights of the connections. Many aspects of brain function, especially the learning process, are closely associated with the adjustment of these connection strengths. Brain activity is represented by particular patterns of firing

[23] These are essentially the transmission lines of the nervous system. They are microscopic in diameter and conduct electrical impulses.

[24] Dendrites are in effect the cell body and extend like the branches of a tree. The origin of the word dendrite is from the Greek word for tree.

activity amongst the network of neurons. This simultaneous cooperative behavior of a huge number of simple processing units is at the heart of the computational power of the human brain

Artificial neural networks aims to simulate various properties of biological neural networks. These are computers whose architecture is modeled on the brain. They consist of many hundreds of simple processing units which are wired together in a complex communication network. Each unit or node is a simplified model of a real neuron which fires[25] if it receives a sufficiently strong input signal from the other nodes to which it is connected. The strength of these connections may be varied in order for the network to perform different tasks corresponding to different patterns of node firing activity. The objective is to solve a particular problem, and artificial neural networks have been successfully applied to speech recognition problems, image analysis, and so on. Many of the existing artificial neural networks are based on statistical estimation and control theory. Artificial neural networks have also been applied to the cognitive modelling field.

There are similarities between the human brain and a very powerful computer with advanced parallel processing. Artificial neural networks have provided simplified models of the neural processing that takes place in the brain. The challenge for the field is to determine what properties individual neural elements should have to produce something useful representing intelligence.

Neural networks are quite distinct from the traditional von Neumann architecture[26] described in Chapter 2. The latter is based on the sequential execution of machine instructions. The origins of neural networks lie in the attempts to model information processing in biological systems. This relies more on parallel processing as well as on implicit instructions based on pattern recognition from sense perceptions of the external world. The nodes in an artificial neural network are composed of many simple processing units which are connected into a network. Their computational power depends on working together (parallel processing) on any task, and there is no central processing unit following a logical sequence of rules. Computation is related to the dynamic process of node firings. This structure is much closer to the operation of the human brain, and leads to a computer that may be applied to a number of complex tasks.

5.12 Expert Systems

An expert system is a computer system that contains the subject-specific knowledge of one or more human experts. These systems allow knowledge to be stored and intelligently retrieved. Expert Systems arose in Artificial Intelligence during the

---

[25] The firing of a neuron means that it sends off a new signal with a particular strength (or weight).

[26] In fact, many computers today still have an underlying von Neumann architecture. However, there have been major advances in speed and size of the hardware since the 1940s. Computers consist of a Central Processing Unit which executes the software (set of rules), and any intelligence in the machine resides in the set of rules supplied by the programmer.

1960s, and several commercial expert systems have been developed since then. An expert system is a program made up of a set of rules (or knowledge) supplied by the subject-matter experts about a specific class of problems. The success of the expert system is largely dependent on the quality of the rules provided by the expert. The expert system also provides a problem solving component that allows analysis of the problem to take place, as well as recommending an appropriate course of action to solve the problem. That is, the expert system employs a reasoning capability to draw conclusions from facts that it knows, and it also recommends an appropriate course of action to the user.

Expert Systems have been a major success story in the AI field. They have been employed in various areas such as for diagnostics in the medical field, equipment repair, and investment analysis. There is a clear separation between the collection of knowledge and the problem solving strategy of the system. Expert systems consist of the following components (Table 5.3):

Table 5.3 Expert systems

Component	Description
Knowledge Base	The knowledge base is composed by experts in a particular field, and the expert system is only as good as the knowledge provided by the experts.
Communication Components	This component (typically a user interface) allows the user to communicate with the expert system. It allows the user to pose questions to the system and to receive answers (or vice versa).
Context Sensitive Help	This is usually provided on all screens and the help provided is appropriate to where the user is in the system.
Problem Solving Component	The problem solving component analyses the problem and uses its inference engine to deduce the appropriate course of action to resolve the problem. The inference engine is software that interacts with the user and the rules in the knowledge base to produce an appropriate solution.

Expertise consists of knowledge about a particular domain and involves understanding problems in the domain as well as skill at solving these problems. Human knowledge of a specialty is of two types: namely public knowledge and private knowledge. The former includes the facts and theories documented in text books and publications on the domain, whereas the latter refers to knowledge that the expert possesses that has not found its way into the public domain. The latter often consists of rules of thumb or heuristics that allow the expert to make an educated guess where required, as well as allowing the expert to deal effectively with incomplete or erroneous data. The challenge for the AI field is to allow the representation of both public and private knowledge to enable the expert system to draw valid inferences.

The inference engine is made up of many inference rules that are used by the engine to draw conclusions. Rules may be added or deleted without affecting other rules, and this reflects the normal updating of human knowledge. That is, out-dated facts are, in effect, deleted and are no longer used in reasoning, while new knowledge is applied in drawing conclusions. The inference rules use reasoning which is closer to human reasoning. There are two main types of reasoning with inference rules and these are backward chaining and forward chaining. Forward chaining starts with the data available, and uses the inference rules to draw intermediate conclusions until a desired goal is reached. Backward chaining starts with a set of goals and works backwards to determine if one of the goals can be met with the data that is available.

The expert system makes its expertise available to decision makers who need answers quickly. This is extremely useful as often there is a shortage of experts, and the availability of an expert computer with in-depth knowledge of specific subjects is therefore very attractive. Expert systems may also assist managers with long-term planning. There are many small expert systems available that are quite effective in a narrow domain. The long-term goal is to develop expert systems with a broader range of knowledge. Expert systems have enhanced productivity in business and engineering, and there are several commercial software packages available to assist.

Some of the well-known expert systems that have been developed include Mycin, Colossus and Dendral. Mycin was developed at Stanford University in the 1970s, and it was developed from the original Dendral system. It was written in LISP and it was designed to diagnose infectious blood diseases, and to recommend appropriate antibiotics and dosage amounts corresponding to the patient's body weight. Mycin had a knowledge base of approximately 500 rules and had a fairly simple inference engine. Its approach was to query the physician running the program with a long list of yes/no questions. Its output consisted of various possible bacteria that could correspond to the blood disease, along with an associated probability that indicated the confidence in the diagnosis, and it also included the rationale for the diagnosis and a course of treatment appropriate to the diagnosis.

Mycin's performance was reasonable as it had a correct diagnosis rate of 65%. This was better than the diagnosis of most physicians who did not specialise in bacterial infections. However, its diagnosis rate was less than that of experts in bacterial infections who had a success rate of 80%. Mycin was never actually used in practice due to legal and ethical reasons in the use of expert systems in medicine. For example, if the machine makes the wrong diagnosis who is to be held responsible?

Colossus was an expert system used by several major Australian insurance companies. It helps insurance adjusters to assess personal injury claims, and helps to improve consistency in the claims process. It aims to make claims fair and objective by guiding the adjuster through an objective evaluation of medical treatment options, the degree of pain and suffering of the claimant, and the extent that there is permanent impairment to the claimant, as well as the impact of the injury on the claimant's lifestyle. It was developed by Computer Sciences Corporation (CSC).

Dendral (Dendritic Algorithm) was developed at Stanford University in the mid-1960s, and it was the first use of Artificial Intelligence in medical research. Its objectives were to assist the organic chemist with the problem of identifying unknown organic compounds and molecules by computerized spectrometry. This involved the analysis of information from mass spectrometry graphs and knowledge of chemistry. Dendral automated the decision-making and problem-solving process used by organic chemists to identify complex unknown organic molecules. It was written in LISP and it showed how an expert system could employ rules, heuristics and judgment to guide scientists in their work.

5.13 Review Questions

1. Discuss Descartes and his Rationalist philosophy and his relevance to Artificial Intelligence.
2. Discuss the Turing Test and its relevance on Strong AI. Discuss Searle's Chinese Room rebuttal arguments. What are you own views on Strong AI?
3. Discuss the philosophical problems underlying Artificial Intelligence.
4. Discuss the applicability of Logic to Artificial Intelligence.
5. Discuss Neural Networks and their applicability to Artificial Intelligence.
6. Discuss Expert Systems.

5.14 Summary

The origin of the term "Artificial Intelligence" was a proposal for the Dartmouth Summer Research Project on Artificial Intelligence in the mid-1950s. The ultimate goal of Artificial Intelligence is to create a thinking machine that is intelligent, has consciousness, has the ability to learn, has free will, and is ethical. It is a multi-disciplinary field, and its branches include:

- Computing
- Logic and Philosophy
- Psychology
- Linguistics
- Machine Vision
- Computability
- Epistemology and Knowledge representation

Turing believed that machine intelligence was achievable and he devised the famous "Turing Test" to judge whether a machine was conscious and intelligent. Searle's Chinese Room argument is a rebuttal that aims to demonstrate that a machine will never have the same cognitive qualities as a human, and that even if a machine passes the Turing Test it still lacks intelligence and consciousness.

McCarthy's approach to Artificial Intelligence is to use logic to describe the manner in which intelligent machines or people behave. His philosophy is that common sense knowledge and reasoning can be formalised with logic. That is, human-level intelligence may be achieved with a logic-based system.

Cognitive psychology is concerned with cognition and some of its research areas include perception, memory, learning, thinking, and logic and problem solving. Linguistics is the scientific study of language and includes the study of syntax and semantics.

Artificial neural networks aims to simulate various properties of biological neural networks. These are computers whose architecture is modeled on the brain. They consist of many hundreds of simple processing units which are wired together in a complex communication network. Each unit or node is a simplified model of a real neuron which fires if it receives a sufficiently strong input signal from the other nodes to which it is connected. The strength of these connections may be varied in order for the network to perform different tasks corresponding to different patterns of node firing activity. The objective is to solve a particular problem, and artificial neural networks have been successfully applied to speech recognition problems and image analysis.

An expert system is a computer system that allows knowledge to be stored and intelligently retrieved. It is a program that is made up of a set of rules (or knowledge). These rules are generally supplied by the subject-matter experts about a specific class of problems. The expert system also provides a problem solving component that allows analysis of the problem to take place, as well as recommending an appropiate course of action to solve the problem. Expert Systems have been a major success story in the AI field.