Lecture 19-20-21

ARTIFICIAL INTELLIGENCE (AI) AND MIND

Overview

The object of research in artificial intelligence (AI) is to discover how to program a computer to perform the remarkable functions that make up human intelligence. This work leads not only to increasing use of computers, but also to an enhanced understanding of human cognitive processes which constitute what we mean by ‘intelligence’, and the mechanisms that are required to produce it. What is needed is a deeper understanding of human intelligence and the human mind. In the first section, we will focus on the various definitions of artificial intelligence, and organize it into the four categories which are, systems that think like humans, systems that act like humans, systems that think rationally and systems that act rationally. In the second and third sections, we will explore the field of AI, and the issue of what computers can do, respectively. In the fourth section, we will argue for the computational model of mind. The basic tenet of this thesis is that the brain is just a digital computer and that the mind is a software program. In the fifth section, we will focus on the relation between AI and the functional theory of mind. In the sixth, we will focus on the creativity aspects of artificial intelligence. In seventh and last sections, we will focus the place mind in artificial intelligence and the reductive nature of artificial intelligence in explaining mind.

Key words: Artificial Intelligence, Creativity, Competency, Mind, and Consciousness

ARTIFICIAL INTELLIGENCE (AI) AND MIND

I. What is Artificial Intelligence?

It is difficult to give a precise definition of artificial intelligence. Some recent artificial intelligence scientists have attempted to define artificial intelligence (in short AI) in various ways.

According to Haugeland, artificial intelligence is, "the exciting new effort to make computers think … machines with minds, in the full and literal sense"[1] For Bellman, it is "the automation of activities that we associates with human thinking, activities such as decision

[1] Haugeland, J., (ed.) Artificial Intelligence: The Very Idea, MIT Press, USA, 1985, quoted in Stuart J Russell and Peter Norvig, *Artificial intelligence: A Modern Approach*, Tan Prints (India) Pvt., New Delhi, 2002, p.5.

making, problem solving, learning…"[2] Charniak and McDermett define AI as "the study of mental faculties through the use of computational model."[3] And for Winston, it is "the study of the computations that make it possible to perceive, reason and act."[4] AI, for Kurzweil is "the art of creating machines that perform functions that require intelligence when performed by people."[5] Rich and Knight say that AI is "The study of how to make computers thinks at which, at the movement, people are better."[6] For Schalkoff, AI is "a field of study that seeks to explain and emulate intelligent behavior in terms of computational process."[7] Luger and Stubblefield hold it to be "the branch of computer science that is concerned with the automation of intelligent behaviour."[8]

Let us look at all the definitions from different angles. Haugeland and Bellman point out that artificial intelligence is concerned with thought process and reasoning. They have explained the mind as a machine that is completely associated with human thinking. That is to say, computers do think. But Schalkoff, Luger and Stubblefield are concerned with the behavioural

---

[2] Bellman, R. E., *An Introduction to Artificial Intelligence: Can Computers Think*?, Boyd and Fraser Publishing Company, San Francisco, USA, 1978, quoted by Stuart J Russell and Peter Norvig, in *Artificial intelligence: A Modern Approach*, p.5.

[3] Charniak, Eugene & McDermott, Drew, *Introduction to Artificial Intelligence*, Addission Wesley Publishing Company, Canada 1985, p.6.

[4] Winston, Patrick Henry, *Artificial Intelligence*, Addison-Wesley Publishing Company, London, July 1984, p.2.

[5] Kurzweil, R., The Age of Intelligent Machines, The MIT Press, Massachusetts, 1990, quoted by Stuart J Russell and Peter Norvig, in *Artificial intelligence; A Modern Approach*, p.5.

[6] Rich, Elaine, *Artificial Intelligence*, McGraw-Hill, Inc., Singapore, 1984, p.1.

[7] Schalkoff, R. J., Artificial Intelligence: An Engineering Approach, McGraw-Hill, New York, 1990, quoted by Stuart J Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, p. 5.

[8] Luger, G. F. and Stubblefield, W. A., *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, Benjamin/Cummings, California, 1993, p.2.

aspects of systems. For them, computers behave as intelligently as human beings. Moreover, Kurzweil, Rich and Knight are concerned with or measure success in terms of human performance. For them, artificial intelligence can be attributed to machines, but it belongs basically to the human mind. At last, Charniak, McDermett and Winston are concerned with an ideal intelligence. They explain the mental faculties through the use of computational models.

To sum up, all the definitions of AI can be organized into four categories.[9] They are as follows:

(i) Systems that think like humans.
(ii) Systems that act like humans.
(iii) Systems that think rationally.
(iv) Systems that act rationally.

Now, we have to look at each aspect in detail.

(i) Acting Humanly: Turing Machine Approach

The Turing test, as named after Alan Turing, was designed to provide a satisfactory operational definition of intelligence. Turing defined that intelligent behavior as the ability to achieve human-level performance in all cognitive tasks to fool an interrogator. In his 'C*omputing Machinery and Intelligence*', Turing says the new form of the problem can be described in terms of a game which we call the '*imitation game*.' It is played by a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He or She knows them by labels X and Y, and at the end of the game,

[9] Russell, Stuart J. and Norvig, Peter, *Artificial Intelligence: A Modern Approach*, p.9.

he or she says, either ‘X is A and Y is B’ or ‘X is B and Y is A.’ The interrogator is allowed to put questions to A and B.[10]

Thus, C: will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer to the question. It is A’s object in the game to try to cause C to make the wrong identification. His or her answer might therefore, be ‘my hair is singled, and the longest strands are about nine inches long.’

However, because the tones of voice may not help the interrogator, the answer should be written or better still be typewritten. The ideal arrangement is to have a tele-printer for perfect communication. Alternatively, an intermediary can repeat the questions and answers. The object of the game for the second player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add to her answer such things as ‘I am the woman, do not listen to him,’ but it is of no avail as the man can make similar remark.

Now, we can ask the question, what will happen when a machine takes the part of A in this game? Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between man and a woman?

Turing’s answers to these questions are more or less summed up in the following passage; “I believe that in about fifty years time it will be possible to program computers, with a

[10] See, Turing, Alan, “Computing Machinery and Intelligence,” in *Minds and Machines*, A.R. Anderson (ed), Prentice Hall.,Inc., Englewood Cliffs, New Jersy, 1964, pp.4-8.

storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning."[11]

What Turing had predicted at that time, now is a fact that the machine or the computer can imitate human behaviuor. It should be pointed out that Turing's beliefs about the capabilities and capacities of machines are not limited to such activities as playing the imitation game as successfully as human beings. Roughly speaking, the test Turing proposed is that the computer should be interrogated in the place of human beings.

Turing's test deliberately avoided direct physical interaction between the interrogator and the computer, because physical limitation of a person is unnecessary for intelligence. However, the so-called Turing test includes a-video signal so that the interrogator can test the subject's perceptual abilities. In order to pass through total Turing test, the computer will need computer vision to perceive objects and robotics to move them.

Again, the issue of acting like a human comes up primarily when artificial intelligence programs have to interact with people, as when expert system explains how it came to its diagnosis, or a natural language processing system has a dialogue with a user. These programs must behave according to certain normal covertness of human interaction in order to make them understood. The Turing test shows that machines can interact with human beings the way human beings interact amongst themselves. That is to say that machines can behave the way the human beings do.

---

[11] *Ibid*., p.13.

(ii) Thinking Humanly: The Cognitive Modeling Approach

The interdisciplinary field of cognitive science brings together computer models from Artificial Intelligence and experimental techniques from cognitive psychology to try to construct precise and testable theories of the workings of the human mind. And if we are going to say that a given program thinks like a human being, we must have some way of determining how human beings think. For that, we need to get inside the actual workings of the human mind. Stuart Russell and Peter Norvig say, "there are two ways to do this: through introspection–trying to catch our own thoughts as they go by –or through psychological experiments. Once we have a sufficiently precise theory of the mind, it becomes possible to express the theory as a computer program. If the program's input/output and timing behavior matches human behaviour, that is evidence that some of the program's mechanisms may also be operating in humans."[12]

Now it is almost taken for granted by many psychologists that a cognitive theory should be like a computer program. But we know that cognitive science is the science of mind. Therefore cognitive scientists seek to understand perceiving, thinking, remembering, understanding language, learning and other mental phenomenon. Their research is remarkably diverse, ranging from observing children's mental operation, through programming computers to do complex problem solving, to analyze the nature of meaning. In order to appreciate the work in artificial intelligence, which is a necessary part of cognitive science, it is necessary to have some familiarity with theories of human intelligence. The cognitive scientists introduce the notion of machine intelligence and emphasize the relationship between human and machine intelligence. The aim of artificial intelligence is to develop and test computer programs that exhibit characteristic of human intelligence. The most fundamental contribution of symbolic computational modeling has been the physical symbol system hypothesis. According to Newell and Simon, "a physical symbol system has the necessary and sufficient means for general

[12] Russell, Stuart J. and Norvig, Peter, *Artificial Intelligence: A modern Approach,* p. 6.

intelligent action,"[13] By 'necessary' we mean that any system that exhibits general intelligence will prove upon analysis to be a physical symbol system. By 'sufficient' we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence. Lastly, by 'general intelligent action' we wish to indicate the same scope of intelligence as we see in human action; that in any real situation behaviour appropriate to the events of the system and adaptive to the demands of the environment can occur within some limits.

However, the ability of computer simulations to model such process is interpreted as a proof of the broader claim that a symbol system is at the center of human intelligence. In this hypothesis it shows that intelligence is an essential aspect of machines. If the machines have the capacity of intelligence, intelligence is the essence of human cognitions. Therefore, machines have cognitive capacity like the human beings. In the cognitive modeling approach, thus, human beings and machines show the property of being intelligent.

(iii) Thinking Rationally: The Laws of Thought Approach

"Right thinking" is the inferential character of every reasoning process. Aristotle in his famous syllogisms provided patterns of argument structures that always give correct conclusions from given correct premises. In the syllogisms, the Laws of Thought play a vital role because these give law the right explanation of a syllogistic inference. There are three Laws of Thought recognized by the logicians. These have traditionally been called the law of Identity, the law of

[13] Newell, A. and Simon, H. A., "Computer Science as Empirical Science: Symbols and Search," in *Mind Design*: *Philosophy, Psychology, and Artificial Intelligence,* John Haugeland (ed.), The MIT Press, Cambridge, Mass., 1981, p.41.

Contradiction, and the law of Excluded–Middle.[14] These Laws of Thought are appropriate to different contexts. The formulations appropriate as follows:

a) The law of Identity asserts that *if any statement is true, then it is true.* This law asserts that every statement of the form P ⊃ P is true, and that every such statement is a tautology.
b) The law of Contradiction asserts that *no statement can be both true and false*. This law asserts that every statement of the form P . ~P is false, that is, every such statement is self-contradictory, and its negation is logically true.
c) The law of Excluded-Middle asserts that *any statement is either true or false*. This law asserts that every statement of the form P ∨ ~P is true, that is, every such statement is a tautology.

In the 'Laws of Thought' approach to artificial intelligence, the whole emphasis is on correct syllogistic inferences. For example-

"Socrates is a man;

All men are mortal;

Therefore, Socrates is mortal."

In this inference, the conclusion is based on the premises according to the rules of inference. The above syllogistic inference is the best example to formulate an AI program. In all reasoning of this type the emphasis is on the logical inference of a conclusion from the premises. In the AI programme this type of logical inference is of much use since this programme provides a variety of logical reasoning. In an inference, a set of variables, a set of constant terms, a set of functions, the set of connectives *if, and, or,* and, *not,* the quantifiers '*exists'* and '*for all'* are the most important symbols to build an AI program. All these constants and variables are the

[14] Copi, Irving M., and Cohen, Carl, *Introduction to Logic*, Macmillan Publishing Co., Inc, New York, 1982, p.319.

arbitrary representations of the world. With the help of these symbols, the so-called logistic tradition within artificial intelligence hopes to build on such programs to create intelligent systems.

(iv) Acting Rationally: The Rational Agent Approach

Here, the word 'agent' refers to mechanical agent (computer). Acting rationally means acting so as to achieve one's goal, given one's beliefs. An agent (mechanical) is something that perceives and acts. Stuart Russell and Peter Norvig point out that making correct inferences is a part of being a rational agent because one way to act rationally is to reason logically to the conclusion that a given action will achieve one's goals, and then to act on that conclusion.

According to them, "the study of artificial intelligence as rational agent design therefore has two advantages. First, it is more general than the 'laws of thought' approach, because correct inference is only a useful mechanism for achieving rationality, and not a necessary one. Second, it is more amenable to the scientific development than approaches based on human behavior or human thought, because the standard of rationality is clearly defined and completely general."[15] As we have seen, an agent is something that perceives and acts in an environment. The job of artificial intelligence is to design the agent program, that is a function that implements the agent mapping from percepts to action. We assume this program will run on some sort of computing device A human agent has eyes, ears, and other organs for sensors, and hands, legs, mouth and other body parts for effectors. The relationship among agents, architectures, and programs can be summed up as:

Agent = Architecture + Program.

---

[15] Russell, Stuart J. and Norvig, Peter, *Artificial Intelligence: A Modern Approach*, pp. 7-8.

The agent is autonomous to the extent that its behaviour is determined by its own experience. A truly autonomous intelligent agent should be able to operate successfully in a wide variety of environments, given sufficient time and scope. But before we design an agent, we must have a pretty good idea of the possible percepts and actions, the agent's goal is supposed to achieve, and what sort of environment it will operate in it.

Again, as we have mentioned, an agent is just something that perceives and acts. In this approach, artificial intelligence is viewed as the study and construction of a rational agent. One of the important factors is that correct inference is not the whole of rationality. There are also ways of acting rationally that cannot reasonably be said to involve inference. For example, pulling one's hand off a hot stone is a reflex action that is more successful than a slower action taken after careful deliberation.

II. The Field of Artificial Intelligence

Artificial intelligence is a new research area of growing interdisciplinary interest and practical importance. People with widely varying backgrounds and professional knowledge are contributing new ideas and introducing new tools into this discipline. Cognitive minded psychologists have developed new models of the mind based on the fundamental concepts of artificial intelligence, symbols systems and information processing. Linguists are also interested in these basic notions while developing different models in computational linguistics. And philosophers, in considering the progress, problems and potential of this work towards non-human intelligence, have sometimes found solution to the age-old problems of the nature of mind and knowledge.

However, we know that artificial intelligence is a part of computer science in which are designed intelligent systems that exhibit the characteristics we associate with intelligence in human behaviour, understanding language learning, reasoning, problem solving, and so on. It is believed that insights into the nature of the mind can be gained by studying the operation of such systems. Artificial intelligence researchers have invented dozens of programming techniques that support some sort of intelligent behaviour.

Now the question is: Is artificial intelligence a science or an Art? According to Tanimoto, “the activity of developing intelligent computer systems employs both proved mathematical principles, empirical results of studying previous system, and heuristic, pragmatic programming techniques. Information stored in rational data structures can be manipulated by well–studied techniques of computer science such as tree searching algorithms. At the same, experimental or vaguely understood ‘rules of thumb’ for problem solving are often crucial to the success of a system and must be carefully accommodated in intelligent systems.”[16] Thus artificial intelligence is both an art and a science. It is a science because it develops intelligent computer systems by employing proved mathematical principles. It is an art because it designs systems by employing programming techniques. Information stored in relational data structure can be manipulated by well-studied techniques of computer science such as tree searching diagram. Thus, the field of artificial intelligence is fascinating because of this complementing of art and science.

Artificial intelligence research may have impact on science and technology in the following way:

(i) It can solve some difficult problems in chemistry, biology, geology, engineering and medicine.

(ii) It can manipulate robotic devices to perform some useful, repetitive, sensory-motor tasks;

[16] Tanimoto, Steven L., *The Elements of Artificial Intelligence,* Computer Science Press, Inc., Maryland, 1987, p.6-7.

Besides, artificial intelligence researchers investigate different kinds of computation and different ways of describing computation in an effort not just to create intelligent artefacts, but also to understand what intelligence is. According to Charniak and McDermott, their basic tenet is to create computers which think. Thus AI expands the field of intelligent activity of human beings in various ways.

Now the question is: What would the world be like if we had intelligent machines? What would the existence of such machines say about the nature of human beings and their relation to the world around them? These questions have raised profound philosophical issues which will be discussed in due course.

The hypothesis of artificial intelligence and its corollaries are empirical in nature whose truth or falsity is to be determined by experiment and empirical test. The method of testing the results of artificial intelligence is of the following types:

(i) In the narrow sense, artificial intelligence is part of computer science, aimed at exploring the range of tasks over which computers can be programmed to behave intelligently. Thus it is the study of the ways computers can be made to perform cognitive tasks, which generally human beings undertake.

(ii) In the wider sense, artificial intelligence is aimed at programs that simulate the actual processes that human beings undergo in their intelligent behaviour. And these simulation programs are intended as theories describing and explaining human performance. But they are tested by comparing the computer output with the human behavior to determine whether both the result and also the actual behavior of computers and persons are closely similar.

A digital computer is an example of a physical symbol system, a system that is capable of inputting, outputting, storing, etc., following different courses of operation. These systems are capable of producing intelligence depending on the level of mechanical sophistication they are.

The computers with these capabilities behave intelligently like human beings, according to the AI researchers.

One of the important tenets of cognitive science is that it stresses the relationship between human and machine intelligence. It is interested in using artificial intelligence techniques to enlighten us about how human beings do intelligent tasks. For example, in getting a machine to solve geometric or integral calculus problems, we are also interested in learning more about the power and flexibility of the human capacity for problem solving.

Thus, the AI researcher attempts to develop and test computer programs that exhibit characteristics of human intelligence. The AI researcher always works with an artifact, which is only inferentially related to the human mind. As Boden says, "The 'machines' in question are typically digital computers, but artificial intelligence is not the study of computers."[17] But it is the study of intelligence in thought and action. Moreover, as Bonnet would say, "artificial intelligence is the discipline that aims to understand the nature of human intelligence through the construction of computer programs that imitate intelligent behaviour."[18]

Further, Bonnet says, "we cannot define human intelligence in general we can highlight a number of criteria by which it can be judged, such as the ability to make abstractions or generalizations, to draw analogies between different situations and to adopt new ones, to detect and correct mistakes in order to improve future performance, and so on"[19] AI is concerned with

[17] Boden, A. M., "Artificial Intelligence", in *The Oxford Companion to Mind,* Richard L. Gregory, (ed.), Oxford University Press, Oxford and New York, 1987, p. 48.

[18] Bonnet, Alain, *Artificial Intelligence: Promise and Performance*, Prentice-Hall International, New Delhi and London, 1985, p.16.

[19] *Ibid.,* p.16.

the intelligent properties of the computational systems, which perform many intelligent tasks. These tasks can be of very varied nature, such as, the understanding of a spoken or written text in a natural language, playing chess, solving a puzzle, writing a poem, making a medical diagnosis, finding one's way from Hyderabad to Delhi. The AI programme selects such activities include information and reasoning processes which are part of any intelligent system.

Artificial intelligence includes commonsensical tasks, such as understanding English language, recognizing scenes, finding a way to reach an object that is far overhead, heavy and frigid, and making sense of the plot of a mystery novel. In addition, artificial intelligence includes expert tasks, such as diagnosing diseases, designing computer systems, locating mineral deposits, and planning scientific experiments. The techniques that artificial intelligence applies to solve these problems are representation and inference methods for handling the relevant knowledge, and search based problem-solving methods for exploiting that knowledge. Although the tasks with which artificial intelligence is concerned may seem to form a very heterogeneous set, in fact, they are related through their common reliance on techniques for manipulation of knowledge and conducting search which we will discuss in the next section.

III. What Computers Can Do

Within the scientific discipline of artificial intelligence, there are several distinct areas of research, each with its own specific interests, research techniques, and terminology. However, as Avron Barr and Edward A. Feigenbaum point out, artificial intelligence is part of computer science concerned with designing intelligent computer systems, that is, systems that exhibit the characteristics we associate with intelligence in human behavior such as understanding language, learning, reasoning, solving problems and so on.

In artificial intelligence, this specialization includes research on language understanding, automatic programming, and several others. The following discussion of the status of artificial intelligence attempts to show the sub–fields identifying some aspects of intelligent behaviours and indicating the state of relevant research.

Like the different sub–fields of artificial intelligence, the different behaviours discussed here are not at all independent; separating them out is just a convenient way of indicating what current artificial intelligence programs can do.

(i) The Problem Solving

The first big successes in artificial intelligence were programs that could solve puzzles and play games like chess. Techniques like looking ahead several moves and dividing difficult problems into easier sub-problems evolved in the fundamental artificial intelligence techniques of search and problem reduction. Another problem solving program that integrates mathematical formulas symbolically has attained very high level of performance and is being used by scientists and engineers. According to Avron Barr and Edward A. Feigenbaum, human beings often solve a problem by finding a way of thinking about it that makes the solution easy, but so far, artificial intelligence programs must be told how to think about the problems they solve.

Elaine Rich defines problem solving, as that which "chooses the best technique(s) and applies it (them) to the particular problem."[20] Another view is that a goal and a set of means of achieving the goal are called a problem solving. Before an agent can start searching for solutions, it must formulate a goal and then use the goal to formulate a problem. According to Stuart Russell and Peter Norvig,[21] a problem consists of four parts. The initial state, a set of operators, a goal test function, and a path cost function. The environment of the problem is represented by a state space. A path through the state space from the initial state to a goal state is a solution. And a search can be judged on the basis of completeness, optimality, time complexity, and space

[20]Rich, Elaine, *Artificial Intelligence*, p. 25.

[21]Russell, Stuart, and Norvig, Peter, *Artificial Intelligence: Modern Approach*, p. 56.

complexity. There are different search paths for an intelligent agent. The agent has to search which is the minimum path to reach the goal. To solve a problem is the main task of the agent in artificial intelligence.

Moreover, one of the important searches in artificial intelligence is the technique of 'heuristic' search. The word 'heuristic' is derived form the Greek verb 'heuriskein', meaning 'to find' or 'to discover'. Some people use heuristic as the opposite of algorithmic. According to Newell, Shaw and Simon, "process that may solve a given problem, but offers no guarantees of doing so is called a heuristic for that problem."[22]

We know that heuristic techniques dominated early application of artificial intelligence. Heuristic method is still used in problem solving. For example, game playing is also a form of problem solving. Games have engaged the intellectual faculties of the humans. And game playing is also one of the oldest areas of endeavour in artificial intelligence. A chess-playing computer is a proof of a machine doing something though intelligence. Furthermore, the simplicity of the rules and their application in the program implies that it is easy to represent the game as a search through a space of possible game positions.

The initial state, the operators, a terminal test, and a pay-off function can define a game. According to Carpenter and Just,[23] an intelligent machine can follow all these rules and play

---

[22] Newell, A., Shaw, A., and Simon, H. A., "Empirical Explorations with the Logic Theory Machine," quoted by Stuart J Russell and Peter Norvig, in *Artificial Intelligence: A Modern Approach*, p. 94.

[23] Carpenter, Patricia A. and Just, Marcel Adam, "Computational Modeling of High-level Cognition Verses Hypothesis Testing," in *The Nature of Cognition*, Robert J. Sternberg (ed.), The MIT Press, Cambridge, Mass., 1999, 246-249.

more efficiently than human beings. For example, in speed chess, computers have defeated the world champion, Gary Kosparov, in both five-minutes and twenty five minutes games. Such a system would be a significant achievement not just for game playing research, but also for artificial intelligence research in general, because it would be much more likely to be applied to the problem faced by a general intelligent agent. There are so many kinds of search in order to solve a problem. But we are not explaining the entire search. We are mainly concerned with how artificial intelligence programs solve different problems within a few seconds.

(ii) Natural Language Processing

The most common way that people communicate is by speaking or writing in one of the natural languages like, Hindi, English, French or Chinese. However, the computer-programming language seems to differ from human languages. These 'artificial' languages are designed so that sentences have a rigid format, or syntax, making it easier for compliers to phrase the programs and convert them into the proper sequences of computer instructions. Besides, being structurally simpler than natural language, programming language can easily express only those concepts that are important in programming: 'Do this, then do that', 'See whether such and such is true'.

One of the important facts is that if computers could understand what people mean when people use English language, the systems would be easier to use and would fit more naturally into people's lives. Artificial intelligence researchers hope that learning how to build computers that can communicate as people do would extend our understanding of human language and mind.

The goal of cognitive linguistics is to specify a theory of language comprehension and production to such a level of detail that a person could write a computer program that could understand and produce natural language. Thus, the intelligent machine program has the capacity to understand natural language. It also responds to questions in an appropriate manner. Thus, if the question were 'Is Rome the capital of France?' the simple reply 'No' would be appropriate but, 'no, it is Paris,' or, 'No, Rome is the capital of Italy' would be more complex. Machine programs would also be able to explain the meaning of the word in other terms and also to translate from one language to another.

(iii) Machine Vision

Vision is the information processing of understanding a scene from its projected images. An image is a two-dimensional function (X, Y) obtained, with a sensing device that records the value of an image feature at all points (X, Y). Images are converted into a digital form for processing with a computer. The task of a computer-vision system is to understand the scene that an imaginary of pixels depicts.

Generally, machine vision means the vision of a machine that can see or perceive the things in the world. According to David Marr, vision is the process of discovering form images what is present in the world, and where it is. Thus, vision is, first and foremost, an information-processing task. Moreover, there is a distinction between human vision (perception) and machine vision, because human vision is defined as "the eye is just a sensor; the visual cortex of the human brain is our primary organ of vision."[24] In the case of human vision, the sensory organ necessarily forms a representation or physical encoding of the received information that facilitates the answering of some questions about the environment, but makes it extremely difficult to answer others. For example, an examination of the encoded information produced by the human eye reveals that at any instant of time the eye has sensed only a small part of the electromagnetic spectrum, and has extracted an image of the scene from a particular viewpoint in

[24] Fischler, Martin A., and Firschein, Oscar, *Intelligence: The Eye, The Brain and The Computer*, Addision-Wesley Publishing Company, Mass., 1987, p.208.

space. It possesses the visual information with the help of the lens fitted within it. Thus machine vision works more or less like human vision.

Learning is now perceived as a gateway to understanding the problem of intelligence. Because seeing is also intelligence, seeing is also learning which is a key to the study of intelligence and biological vision. Visual neuroscience develops the methods of understanding how human visual systems work. Visual neuroscience is beginning to focus on the mechanisms that allow the cortex to adapt its circuitry and learn a new task. Machine vision as a part of the AI programme, however, tries to develop systems that can simulate human vision.

(iv) Machine Learning

Learning for an intelligent agent is essential for dealing with unfamiliar environments. Learning a function from examples of its input and output is called inductive reasoning. According to Bonnet, the ability to learn is one of the fundamental constituents of intelligence, 'learning' being understood in its general sense as indicating the way in which the humans and animals can increase their stock of knowledge and improve their skill and reasoning powers.

Moreover, from the very beginning of artificial intelligence, researchers have sought to understand the process of learning and to create computer programs that show learning behaviour. Ordinarily, learning is a very general term denoting the way in which people increase their knowledge and improve their skills. There are two fundamental reasons for studying learning. One is to understand the process itself. By developing computer models of learning, psychologists have attempted to gain an understanding of the way humans learn. Philosophers since Plato have also been interested in the process of learning, because it might help them understand and know what knowledge is and how it grows.

The second reason for learning research is to provide computers with the ability to learn. Learning research has potential for extending the range of problems to which computers can be applied. In particular, the work in machine learning is important for expert systems development, problem solving, computer vision, speech understating, conceptual analysis of databases, and intelligent authoring systems.

In this way, a computer can do more work than human beings. According to James R. Slagle, computers can do arithmetic at an unbelievable speed, and no human can compete with them. Besides, computers can do many other odd jobs like detecting a certain disease from the available data. Thus, AI as a study of machine intelligence, especially of computers and robots, opens up new vistas of understanding mind and intelligence. Intelligent behaviour can be studied in many dimensions. This is made possible by the invention of computers as high-processing computing machines. The computers can do many things that seem to require intelligence and other mental abilities.

IV. The Computational Model of Mind

As we have already seen, artificial intelligence is the discipline that aims to understand the nature of human intelligence through the construction of computer programs that imitate intelligent behavior. It also emphasizes the functions of the human brain and the analogical functioning of the digital computer. According to one extreme view, the human brain is just a digital computer and the mind is a computer program. This view, as John Searle calls it is strong artificial intelligence.

According to strong artificial intelligence, "the appropriately programmed computer with the right inputs and outputs literally has a mind in exactly the same sense that you and I do."[25] This tells that not only the devices would just referred to indeed be intelligent and have minds, etc. but mental qualities of a sort can be attributed to teleological functioning of any computational device, even the very simplest mechanical ones such as a thermostat. Here, the idea is that mental activity is simply the carrying out of some well-defined operations, frequently referred as an algorithm. We may ask here as to what an algorithm actually is. It will be adequate to define an algorithm simply as a calculation procedure of some kind. But in the case of thermostat, the algorithm is extremely simple: the device registers whether the temperature is greater or smaller than the setting, and then, it arranges for the circuit to be disconnected in the former case and to remain connected in the latter. For understanding any significant kind of mental activity of a human brain, a very complex set of algorithms has to be designed to capture the complexity of the human mental activities. The digital computers are approximations to the complex human brain.

The strong artificial intelligence view is that the differences between the essential functioning of a human being (including all its conscious manifestations) and that of a computer lies only in the much greater complication in the case of brain. All mental qualities such as thinking, feeling, intelligence, etc. are to be regarded, according to this view, merely as aspects of this complicated functioning of the brain; that is to say that they are the features of the algorithm being carried out by the brain. The brain functions like a digital computer according to this view.

The supporters of strong AI hold that the human brain functions like a Turing machine which carries out all sets of complicated computations. The brain is naturally designed like a computing machine to think, calculate and carry out algorithmic activities. To strong AI

[25] Searle, John R., "Minds and Brains without Programs," in *Mindwaves: Thoughts on Intelligence, Identity and Consciousness*, C. Blakmore, and S. Greenfield (ed.), Bail Blackwell, Oxford, 1987, p.210.

supporters, the activities of the brain are simply algorithmic activities which give rise to all mental phenomena like thinking, feeling, willing, etc.

David Chalmers, mentions that the field of artificial intelligence is devoted in large part to the goal of reproducing mentality in computational machines. The supporters of strong AI argue that we have every reason to believe that eventually computers will truly have minds. Winston says that, "intelligent robots must sense, move and reason."[26] Accordingly, intelligent behaviour is interpreted as giving rise to abstract automation. That is to say that an artificial, non-biological system could thus be the sort of thing that could give rise to conscious experience. For the supporters of strong AI, humans are indeed machines, and in particular, our mental behaviour is finally the result of the mechanical activities of the brain.

John Searle, in his "*Is the Brain a digital computer*?"[27] mentions that the basic idea of the computer model of the mind is that the mind is the software and the brain is the hardware of a computational system. The slogan is: "the mind is to the program, the brain is to the hardware."[28] For strong AI, there is no distinction between brain processes and mental processes. Because the process which is a happening in the brain is a computational process, the mind is the alternative name of the brain which is a machine.

According to David Chalmers, the theory of computation deals wholly with abstract objects such as Turing machine, Pascal program, finite state automation and so on. These abstract objects are formal structures which are implemented in formal systems. However, the

[26] Winston, P. H., *Artificial Intelligence,* p.380.

[27] Searle, John R, '*Is the Brain a digital computer*?", received a paper from internet.(www.ecs.sotoh.ac.uk/~harnad/papers/py104/searle.comp.html.)

[28] *Ibid*.

notion of implementation is the relation between abstract computational objects and physical systems. Computations are often implemented in synthetic silicon based computers.

Whereas the computational systems are abstract objects with a formal structure determined by their states and state transition relations, the physical systems are concrete objects with a causal structure determined by their internal states and the causal relations between the states. It may be pointed out that a physical system implements a computation when the casual structure of the system mirrors the formal structure of the computation. The system implements the computation, if there is a way of mapping states of the system onto states of the computations so that the physical states which are causally related map onto the formal stares that are correspondingly formally related.

The fact is that there is rich causal dynamics inside computers, as there is in the brain. There is real causation going on between various units of brain activity precisely mirroring patterns of causation between the neurons. For each neuron, there is a specific causal link with other neurons. It is the causal patterns among the neurons in the brain that are responsible for any conscious experiences that may arise.

The brain, as Marvin Minsky says, "happens to be a meat machine."[29] He points out that the brain is an electrical and chemical mechanism, whose organization is enormously complex, whose evaluation is barely understood and which produces complex behavior in response to even more complex environment. Artificial intelligence understands the nature of human intelligence in terms of the computational model of the brain.

[29] This view quoted by Pamela McCorduck, in his *Machines Who Thinks*, W.H. Freeman and Company, San Francisco, 1979, p., 70.

V. Artificial Intelligence and the Functionalist Theory of Mind

Functionalism arose as a result of the phenomenal rise of interest in computing machines and artificial intelligence. The functionalists say that mental processes are computational processes realized in a machine. Functionalism is a theory that explains mental phenomenona in terms of the external input and the observable output. It explains the mind as a machine.

Functionalism as a conception of mind explains the reality of the mental phenomena in terms of the mechanical phenomena. Mind is not just a part of the physical organism but is itself characterized as the physical organism called the brain. The brain is further characterized as a complex physical system working like a computing machine. Functionalism explains the mechanical behaviour of the brain/mind.

Functionalism as a theory of mind is supported by various scientific theories like those of artificial intelligence, cognitive science, and neuroscience, etc., Artificial intelligence advocates a computational theory of mind which argues in favour of the functional similarity between the computation states of the artificial system and the neurophysiological states of the brain.

The hypothesis of artificial intelligence that "machine can think" became very popular after Alan Turing's "C*omputing Machine and Intelligence.*" Turing's hypothesis is that machines think intelligently like human beings. Putnam writes, "Probabilistic automation has been generalized to allow for 'sensory inputs', and 'motor outputs', -that is, the Machine Table specifies, for every possible combination of 'state' and a complete state of 'sensory inputs' an 'instruction' which determines the probability of the next 'state' and also probabilities of the ‘motor out puts’.”[30] The following are the steps which explain how a machine functions in general:

(i) The description of the sequence of states (procedure)
(ii) Description of rules
(iii) The explanation of the rationale of the entire procedure.

The computing machine is thus a system constructed out of different subsystems that function inside it to process the inputs and to produce the output once the input is simulated in the machine. It tries to match the simulating state with the states already computed and mapped in the system. This mapping order follows certain syntax or rules. The syntax is responsible for the correlation of total cognitive states. Thus, the entire process of simulation can be called an intelligent process. This simulation process takes place between the functions of the two functionally isomorphic systems. As Putnam defines, “Two systems are functionally isomorphic if there is a correspondence between states of one and the states of the other that preserves the functional relation.”[31]

There is functional isomorphism, according to Putnam, between the brain/mind and a machine. This functional isomorphism holds due to the causal capacity of the functional states of the machine. For example, when I have a pain, there is a neurophysiological process corresponding to the mental state because of the firing of the C-fiber. The brain/mind identity

[30] Putnam, H., “The Nature of Mental States”, *The Philosophy of Mind: Classical Problem/ Contemporary Issues*, Brain Beakley and Peter Ludlow (ed.), The MIT Press, Cambridge, Mass., 1992, p. 54.

[31] Putnam, H., *Mind, Language and Reality: Philosophical Papers 2*, Cambridge University Press, Cambridge, 1976, p. 291.

follows as there is functional identity between the two. Thus, identity between the mental states and the physical processes of the brain is established from the functional point of view. That is, in functional terms, the brain state is isomorphic with the mental state. That is, there is identity between software that constitutes the program and the hardware of the machine, which helps the software to be realized in the machine.

There can be indefinitely many different physical properties, which constitute the realizations of the same functional property. However, it is also true that the same physical state can realize different functional properties at different times or in different circumstances or in different creatures. The functional states are 'multiply realizable' in the sense that a functional state cannot be identical to any particular physical realization of it. For example, someone could write a program using two completely different types of computer, which use different sorts of hardware to run the same program. In this sense, the program is said to be 'multiply realizable' in that any number of computers may be used to realize the same program. Functionalism takes states of mind and mental properties to be functional states and properties. Mental properties are realizable by, but not identical with, material properties. For example, the same mental property, the property of being in pain, may be realized by one property in a human being and to a certain extent by another property in an invertebrate. For the functionalist, if someone has now a particular pain, then he/she can imagine that this pain is realized through a particular neural state. That neural state has an identifiable material structure, and this may be studied by a lower-level hardware science like neurobiology. Therefore, for functionalism, what makes the state a realization of pain, is not its material constitution but it's occupying a particular kind of causal role within our nervous system. Multiple realizability thus implies that there is a higher-level functional description of physical states in terms of their causal role, which abstracts from their lower-level physical constitution. It is with such functional properties that mental properties can be identified.

In his essay "*Mad Pain and Martian Pain*",[32] David Lewis discusses two kinds of beings, which experience pain differently than normal humans. In the case of mad pain, the subject experiences pain when doing moderate exercise in an empty stomach; further, it improves his concentration for mathematical reasoning. On the other hand, Martian pain takes place in a Martian organism constructed of hydraulic hardware rather than neurons. Here the point is that pain is associated only contingently with either its causes (as in mad pain) or its physical realization (as in Martian pain). We cannot specify *a priori* its causal role or physical

Daniel Dennett has suggested a multiple-draft-model, approach to the nature of mind. According to this model, there is similarity between the functions of the human mind and those of the computer. The brain system functions in relation to different sub-systems. So there are multiple drafts, which operate within an artificial system. Such an analogy is beneficial because it analyses consciousness from the point of view of language processing. This is given importance precisely in the sense that a linguistic or language speaking being is considered not only as a conscious being but also a rational being. Even the robots as information processing systems can also be characterized as intelligent systems. According Dennett, "of course, we are machines! we are just very, very complicated, evolved machines made of organizes molecules instead of metal and silicon, and we are conscious, so there can be conscious machines – us."[33] So the human thought process and language processing in the artificial systems are analogous to each other. In the case of the conscious thought process, we are aware of our thoughts, at the same time, there is physico-chemical process, which goes on in our brain.

Dennett's functional analysis of consciousness is divided into two parts. There are the sub-personal view of consciousness and the multiple draft-model of consciousness respectively.

---

[32] Lewis, David, *Philosophical Papers*, Volume 1, Oxford University Press, Oxford, 1983, p. 122.

[33] Dennett, D.C., *Consciousness Explained* Penguin Books, New York and London, 1991, p. 431.

The sub-personal model explains consciousness and other mental activities through the help of neurological states and processes of the organism, whereas the multiple-draft-model discusses how an artificial system behaves intelligently. Thus Dennett provides a functional explanation of consciousness at the sub-personal level. According to him, “Sub-personal theories proceed by analyzing a person into an organization of subsystems (organs, routines, nerves, faculties, components-even atoms) and attempting to explain the behaviour of the whole person as the outcome of the interaction of these subsystems. Thus in the present instance the short-coming emerged because the two access notions introduced computational access *simpliciter* and the computational access of a print-out faculty, were defined at the sub-personal level; if introduced into a psychological theory they would characterize relations not between a person and a body, or a person and a state of affairs or a person and anything at all, but rather, at best relations between parts of person (or there bodies) and other things.”[34]

Therefore, the sub-personal level of explanation of consciousness tries to explain not only how the human beings are systems of organism but also how the system is being constituted and how the various functions involved in different physiological parts of the organism function together. And that functional structure would help us in defining the capacity involved in causing consciousness or what we call conscious behaviour. A state of consciousness is simply one which exhibits a certain characteristic pattern of causal relations to other states, both mental and physical.

We know that human beings perform various activities; they learn language; acquire various knowledge states or belief states, and there are changes in their belief states, and so on. All these activities are independent biological activities of human life. Dennett anticipates that there would be a system whose programs would be such that it would be self-dependent in all its functions. That would be able to replace or stand parallel to human intelligence. Further, the

[34] Dennett, D. C., *Brainstorms: Philosophical Essays in mind and Psychology*, The MIT Press, Cambridge, Mass., 1981, p. 153.

functions of the artificial system help us in explaining the various mysterious features that are ascribed the human life. Thus, the strong notion of functionalism advocates the identity between the mental states and the brain processes. It also explains the different basic features of human beings such as consciousness, intentionality and subjectivity, etc, by bringing the feature of the functional isomorphism into account.

Functionalism holds that the mental states are abstract, functional states, characterized solely in terms of their casual relationships to each other, to input, and to outoput. Human purposive behaviour is then explained in terms of how this hypothesized system of states take the organism from sensory input to behaviour. Because functionalism insists upon a network of mental states, it insists upon the holism of the mental upon the way in which mental states operate together to explain behaviour. It accepts structure of mental states in which each is necessarily connected with the other. The mental states do not function in isolation; rather they function within the causal co-relationship with other mental states. The function of mental states also takes into account the effect of the environment in which the subject or agent is placed with the system that must be well equipped to receive the input from the environment and to produce the output.

The functionalist program has been strongly influenced by analogies drawn from computer science and AI, both in its general outlook and in several of its specific applications to problems about the nature of mind. Because a functional state is like a computational state of a computer. A computer program can be described as a functional organization of the hardware. As already discussed, the functionalists argue that mental states are like the 'information processing' states of a computer. Accordingly, for computer functionalism or artificial intelligence, the brain is a computer, and the mind is a computer program implemented in the brain. Thus artificial intelligence is strongly founded on a functionalist conception of mind. It is dependent on the idea that human functions like a digital computer with multifunction computational abilities.

VI. Machine Intelligence, Competence, and Creativity

As already discussed, machine intelligence or artificial intelligence is a physical symbol system, which has the necessary and sufficient means for general action. The above statement shows that machine intelligence has the capacity of competence and creativity. However, the system that exhibits general intelligence will prove upon analysis to be a physical symbol system. Further the physical symbol system of sufficient size can be organized further to exhibit general intelligence. The general intelligent action indicates the same scope of intelligence as we have seen in human action. The symbol system hypothesis implies that intelligence will be realized by a universal computer. It also asserts that the intelligent machine is a symbol system. When machines acquired enough memory to make it practicable to locate actual symbol system with the help of programs produce intelligent behaviour and solve many problems.

Moreover, many machine intelligence scientists use computers as tools, to help them create things they could not have created. For example, many scientists use sounds which no orchestra could produce. A visual artist may get ideas from computer graphics. We are concerned with those programs which produce aesthetically interesting creations. There are a number of programs which explore artistically interesting spaces, and a few which produce aesthetically acceptable results.

Boden has given the example of a famous painter, Harold Cohen. Cohen has written a series of programs which produce pleasing, and unpredictable line drawings. Cohen's programs explore a certain style of line drawing and a certain subject matter. As human artists have to know about the things they are depicting, so each of Cohen's programs needs an internal model of its subject matter. This model is not a physical object. It is a set of abstract rules, which specify not only the anatomy of the human body, but also how the various body-parts appear from various points of view.

AI or Machine intelligence programs that write stories are woefully inadequate compared with human storyteller. But the best of them get what strength they possess from their internal models of very general aspects of motivation. For example, a program is asked to write a story with the moral 'Never trust flatterers.' In this story, there are two characters, the Fox and Crow. The story as follows:

"Once upon a time, there was a dishonest Fox, named, Henry who lived in a cave, and a vain and trusting Crow named Joe who lived in an elm-tree. Joe had gotten a piece of cheese and was holding it in his mouth. One day, Henry walked from his cave, across the meadow to the elm-tree. He saw Joe Crow and the cheese and became hungry. He decide that he might get the cheese if Joe Crow spoke, so he told Joe that he liked his singing very much and wanted to hear him sing. Joe was very pleased with Henry and began to sing. The cheese fell out of his mouth, down to the ground. Henry picked up the cheese and told Joe Crow that he was stupid. Joe was angry, and did not trust Henry any more. Henry returned to his cave. The end."[35]

This program can construct hierarchical plans, ascribing them to the individual characters; according to the sorts of motivation one would expect them to have. It can give one character a role in another's plan. But these roles need not be allocated randomly, but can depend on background interpersonal relations. And it can represent different sorts of communication between the characters, which constrain what follows in different ways. All these matters are represented as abstract schemata, which are used to produce the story-structure. Thus the above programs shows that machines have the capacity to produce creativity. Now the question may be

---

[35] Boden, Margaret A., "Creativity and Computers," in *Artificial Intelligence and Creativity*, Terry Dartnall (ed.), Kluwer Academic Publishers, London and Boston, 1994, p. 14.

raised as; whether programs have scientific discovery? According to Langley,[36] programs designed can find a simple mathematical and classificatory relation as 'rediscovered' with the help of physical and chemical laws.

NET talk is another famous and important example of machine creativity and competence. As we know, this model's goal is successfully to negotiate the problem domain of text-to-speech transformation. NET talk took seven letter segments of text as input and mapped the target window of that input onto an output which coded for phonemes. A connectionists system is not trained but programmed. Let us raise a question: What does NET talk know? According to Clark, "Net talk does not in any sense understand what it is saying. But this is not the point. Likewise, I might learn roughly how to pronounce Chinese sequences without understanding them. Nonetheless, NET talk has gone from babble to an output which is lawfully disciplined with respect to its input. That strongly suggests that it has learnt something. The question, what?"[37]

According to Lenat, automatic mathematician is a system that does not produce proofs, nor solve mathematical problems. Rather, it generates and explains mathematical ideas. Artificial machine starts with 100 very primitive mathematical concepts drawn from set theory, including sets, lists, equality and operations. These concepts are so basic that they do not even include the

---

[36] Langley, P, Simon, H. A., Bradshaw, G. L., and Zytkow, J. M., *Scientific Discovery: Computational Explorations of the Creative Processes*, MIT Press, USA, 1987, quoted by Margaret A. Boden, "Creativity and Computers," in *Artificial Intelligence and Creativity*, Terry Dartnall (ed.), p. 18.

[37] Clark, Andy, "Connectionism and Cognitive Flexibility," in *Artificial Intelligence and Creativity*, Terry Dartnall (ed.), p.65.

ideas of eliminating arithmetic. To begin with, the program does not know what an integer is, still less addition, subtraction, multiplication, and division.

Artificial machine hunches, like human hunches, are sometimes wrong. Nevertheless, it has come up with some extremely powerful notions. It produced many arithmetical concepts, including integer, square root, addition, and multiplication. It generates the fundamental theorem of number theory, though did not prove. And it suggested the interesting idea that every even number greater than two is the sum of two different primes. It has originated one major theorem which no one had ever thought of before. Here, artificial machine appears to be P-creative. According to Boden,[38] there are two senses of creativity. One is psychological creativity (P-creativity) and the other is historical creativity (H-creativity). An idea is P-creative if the person in whose mind it arises could not have had it before; it does not matter how many times other people have already had the same idea. On the other hand, an idea is H-creative if it is P-creative and no one else has ever had it before. Thus artificial machine's creative is P-creative.

Hofstadter presented another important example. He designed a copycat; in this, a program can generate many different analogies, where contextually appropriate comparisons are favoured over inappropriate ones. It does not rely on ready-made, fixed, representations, but constructs its own representations in a context-sensitive way. Its new analogies and new perceptions develop together.

As we have seen, cognitive science tries to provide computational models of the mind, that is, computational simulations of human cognitive process. If creativity is not a computational process, it might still be possible to simulate it computationally, just as it is

---

[38] Clark, Andy, "Connectionism and Cognitive Flexibility," in *Artificial Intelligence and Creativity*, Terry Dartnall (ed.), p.65.

possible to simulate digestive process without the simulation itself being a digestive process. It might be possible to have machine models of human creative processes, even if machines themselves cannot be creative. As Dartnall said, "if machines cannot be creative then I doubt there is any significant sense in which they can be intelligent, for they will never 'have minds of their own.' I do mean this in the weak sense that they will always slavishly do what we tell them, but in a strong-sense, they will never be able to generate their own ideas. And I take it as axiomatic that if they cannot generate their own ideas they cannot be intelligent."[39]

VII. Mind in Artificial Intelligence

This section will explore the states of mind in artificial intelligence. As we know the main aim of artificial intelligence is to reproduce mentality in machines. That is to say that AI aims at producing machines with mind. If we say that machines have minds, then we have to ascribe certain 'belief', 'knowledge', 'free will', 'intention', 'observations', etc. to a machine. In that case, the machines will perform intelligent tasks and thus will behave like human beings.

We may raise a question: Why should we want ascribe mental qualities to machines at all? According to MacCarthy,[40] there are many reasons for ascribing belief and other mental qualities:

[39] Boden, Margaret A., "What Is Creativity," in *Dimensions of Creativity*, Margarent A. Boden (ed.), The MIT Press, Cambridge, Mass., 1994, p.76.

[40] McCarthy, John, "Ascribing Mental Qualities to Machines," in *Philosophical Perspective in Artificial Intelligence*, Martin Ringle (ed.), The Harvester Press, Sussex, 1979, p.164.

(i) Although we may know the program, its state at a given moment is usually not directly observable, and we can explain performance of a machine only by ascribing beliefs and goals to it.

(ii) Ascribing beliefs may allow the derivation of general statements about the machine's behaviour that could not be obtained from any finite number of simulations.

(iii) The difference between this program and another actual or hypothetical program may best be expressed as a difference in belief structure.

According to Haugeland, thought itself is not static and random: It develops in ways that obey different rules of inference. Haugeland says, "since correct application of the rules of reason to particular thoughts depends on what those thoughts mean, it seems that there must be some active rule-applier, which understands the thoughts (and rules), and which applies the rules to the thoughts as well as it can. If the activity of this rules applier, following the rules of reason, is to explain the rationality of our thought process, then it must be regarded as a complete little person – or homunculus (in Latin)- inside the head, directing the thoughts like a traffic cop. The trouble is: a theory that involves an homunculus to explain thinking has begged its own question, because the homunculus itself has to think, and that thinking has not been explained."[41]

Cognitive scientists can be materialists and mentalist at the same time. They are materialist because they support the view that the mind is a complicated machines or matter. On the other hand, some support that along with mind the body exists. They can offer explanation in terms of meaning and rules following, without presupposing any unexplained homunculus. It would be peculiar to start assigning geometrical shapes and locations to the internal program routines and operation of a system. These same decisions clearly cause physical behaviour, yet no one is worried that the laws of physics are being violated. According to Haugeland, "when the

---

[41] Haugeland, John. "Semantic Engines: An Introduction to Mind Design" in *Mind Design: Philosophy, Psychology, Artificial Intelligence,* (ed.) John Haugeland, p.3.

machine plays, it follows rules in at least two senses: it always abides by the rules of the game, and it employs various reasonable rules of thumb to select plausible moves. Though these rules are in no way laws of nature, the machine's behaviour is explained (in part) by citing them- and yet, no unexplained 'compunculus' is presupposed."[42] Thus this explanation will necessary invoke the system's internal reasoning processes; yet it is far from easy to figure out processes that will consistently lead to the observed behavioural response. Dennett rightly says that human mind is a semantic engine, that is to say that the way human mind handles the meaning of a word or sentence, in the same way a machine handles the literal meaning of a word or a sentence. Thus Dennett's view shows that human mind is a machine like ordinary machine because both mind and machine have the same quality, the difference is only apparent.

VIII. Strong AI And Weak AI (Reductionism and Non-Reductionism)

As we have seen, the main thesis of AI is that the human brain is like a digital computer, and the human mind is just a computer program. It tries to prove that the relation between the programs and the computer hardware is like the relation between mind and brain. Some supporters of AI argue that we have every reason to believe that computers have intelligence. At the same time, some others argue that the computers' intelligence is limited whereas human intelligence has no limit. Nowadays computers have achieved some modest success in proving theorems, guiding missiles, sorting mail, driving assembly-line robots, diagnosing illness, predicting weather and economic events, etc. Computers receive, interpret, process, store, manipulate and use information. Thus, intelligent behaviour is programmed into the computers. On the contrary, we have no idea how the brain functions, but we have an idea of the general relationships between brain processes and mental processes. Mental processes are caused by the brain activities which are functions of the elements constituting the brain.

[42] *Ibid.*, p.4.

Strong AI argues that it is possible that one-day a computer will be invented which can function like a mind in the fullest sense of the word. In other words, it can think, reason, imagine, etc., and do all the things that we currently associate with the human minds. On the other hand, weak AI argues that computers can only simulate human mind and are not actually conscious in the same way as human minds are. According to weak AI, computers having artificial intelligence are very powerful instruments in the hands of man. Whereas Strong AI holds that computer is not merely an instrument in the study of the mind, but that the appropriately programmed computer is really a mind in the sense that computers can think and do reasoning like the human beings. In Strong AI, the programmed computer has cognitive states, so the programs are not simple tools that allow us to test psychological explanations; rather the programs are themselves the explanations. Strong AI, according to Searle, basically claims that the appropriately programmed computer literally has cognitive states, and that the programs thereby explain human cognition.

The main aim of AI is to reproduce mentality in computational machines, and to try to prove that the functions of a machine are similar to the functions of the human mind. But the question is: Could a machine have mental states? For AI, in the words of Searle, "there is nothing essentially biological about the human mind. The brain just happens to be one of an indefinitely large number of different kinds of hardware computers that could sustain the programs, which make up human intelligence. On this view, any physical system whatever that had the right program with the right inputs and outputs would have a mind in exactly the same sense that you and I have minds."[43]

Searle is here critical of the view that any physical system that has the right program with the right inputs and outputs would have a mind in exactly the same sense that human beings have minds. The cognitive scientists believe that perhaps they can design the appropriate hardware

[43] Searle, John R., *Minds, Brains and Science*, Harvard University Press, Cambridge, 1984, p.28.

and programs - artificial brains, and minds- that are comparable to human brains and minds. Strong artificial intelligence is a reductionist theory. Because strong AI reduces mind or mentality to physical properties. Here, the term ‘reduces to’ names a relation between theories. When this relation holds between a pair of theories, for example, R^1 and R^2, then R^2 is said to be reducer of R^1. According to Fodor, “the reduction relation is transitive and asymmetrical, hence irreflexible. By the ‘unity of science’ I shall mean the doctrine that all sciences except physics reduce to physics. By ‘physicalistic reduction’ I shall mean a certain claim that is entailed by, but does not entail, the unity of science; viz, the claim that psychology reduces to physics.”[44]

Reducibility involves a good deal more than the ontological claim that things that satisfy descriptions in the vocabulary of R^1 also satisfy descriptions the vocabulary of R^2. This condition is stronger than the ontological requirement that whatever falls under the generalizations of R^1 should also fall under those of R^2. On this view, there is an important sense in which syntax is preserved under reduction. That is to say, reduction permits us to redescribe the events in the vocabulary of R^2. Thus according to strong AI, mental states reduce to the computational states in the same way.

On the other hand, weak AI is non-reductionist, because this theory is not reducing the human mind in terms of machines, but it can only simulate human mind and this does not mean exact replication. The above statement shows that the weak AI view is non-reductionist. For the physicalist, life is a higher order property, which emerges out of the physical properties. However, in case of a zombie, there is an absence of consciousness. In other words, the logical possibility of a zombie world is considered as a world physically identical to our world, but conscious experiences is impossible in this world. The zombies may be psychological or phenomenal zombies, which are physically and functionally identical to human beings but they lack experiences. According to Chalmers, “the logical possibility of zombies seems equally

---

[44] Fodor, J. A., *Representations: Philosophical Essay on the Foundation of Cognitive Science*, p.149.

obvious to me. A zombie is just something physically identical to me but which has no conscious experience – all is dark inside."[45]

The zombie and me have identical physical properties but differ in high-level properties like consciousness. The zombies lack consciousness. Therefore, the high-level property of being conscious cannot be logically supervienent on physical properties. In the same way, according to weak AI, mind and machines have some identical properties but differ mainly on some higher qualities. Here, weak AI is non-reductionist because unlike strong AI, is does not reduce mind to machines.

[45] Chalmers, David J., *The Conscious Mind*, Oxford University Press, New York, 1996, p.96.