Artificial Intelligence[1]

For articles on related subjects *see* ARTIFICIAL LIFE; AUTOMATED PLANNING; COGNITIVE SCIENCE; COMPUTER CHESS; COMPUTER GAMES; COMPUTER MUSIC; COMPUTER VISION; EXPERT SYSTEMS; GENETIC ALGORITHMS; HEURISTIC; KNOWLEDGE REPRESENTATION; LANGUAGE TRANSLATION; MULTI-AGENT SYSTEMS; NATURAL LANGUAGE PROCESSING; NEURAL NETWORKS; PATTERN RECOGNITION; ROBOTICS; SEARCHING; SIMON, HERBERT A.; SPEECH RECOGNITION AND SYNTHESIS; THEOREM PROVING and TURING, ALAN M.

Introduction

Artificial Intelligence[2] (AI) is a field of computer science and engineering concerned with the computational understanding of what is commonly called intelligent behavior, and with the creation of artifacts that exhibit such behavior. This definition may be examined more closely by considering the field from three points of view: computational psychology, computational philosophy, and machine intelligence.

Computational Psychology

The goal of computational psychology is to understand human intelligent behavior by creating computer programs that behave in the same way people do. For this goal it is important that the algorithm expressed by the program be the same algorithm that people actually use, and that the data structures used by the program be the same data structures used by the human mind. The program should do quickly what people do quickly, should do more slowly what people have difficulty doing, and should even tend to make mistakes where people tend to make mistakes. If the program were put into the same experimental situations that humans are subjected to, the program's results should be within the range of human variability.

Computational Philosophy

The goal of computational philosophy is to form a computational understanding of human-level intelligent behavior, without being restricted to the algorithms and data structures that the human mind actually does (or conceivably might) use. By "computational understanding" is meant a model that is expressed as a procedure that is at least implementable (if not actually implemented) on a computer. By "human-level intelligent behavior" is meant behavior that, when

[1] This is a preliminary version of Stuart C. Shapiro, "Artificial Intelligence." In A. Ralston, E. D. Reilly and D. Hemmendinger, Eds. *Encyclopedia of Computer Science, Fourth Edition,* Van Nostrand Reinhold, New York, forthcoming.

[2] This article is a revised version of Shapiro, S. C. "Artificial Intelligence," in S. C. Shapiro, Ed. *Encyclopedia of Artificial Intelligence, Second Edition.* New York: John Wiley & Sons, 1991.

engaged in by people, is commonly taken as being part of human intelligent cognitive behavior. It is acceptable, though not required, if the implemented model perform some tasks better than any person would. Bearing in mind Church's Thesis (*see* CHURCH, ALONZO), this goal might be reworded as asking the question, "Is intelligence a computable function?"

In the AI areas of computer vision (*q.v.*) and robotics (*q.v.*), computational philosophy is sometimes replaced by computational natural philosophy (science). For example, some computer vision researchers are interested in the computational optics question of how the information contained in light waves reflected from an object can be used to reconstruct the object. Notice that this is a different question from the computational psychology question of how the human visual system uses light waves falling on the retina to identify objects in the world, or even the computational philosophy question of how any intelligent entity could use light waves falling on a two-dimensional retinal grid to discriminate one three-dimensional object-in-the-world from a set of other possible objects.

Machine Intelligence

The goal of machine intelligence (called, in earlier versions of this article, "advanced computer science") is to push outwards the frontier of what we know how to program on computers, especially in the direction of tasks that, although we don't know how to program them, people can perform. This goal led to one of the oldest definitions of AI: the attempt to program computers to do what, until recently, only people could do. Although this expresses the idea of pushing out the frontier, it is also perpetually self-defeating in that, as soon as a task is conquered, it no longer falls within the domain of AI. Thus, AI is left with only its failures; its successes become other areas of computer science. The most famous example is the area of symbolic calculus. (*see* COMPUTER ALGEBRA). When James Slagle wrote the SAINT program, it was the first program in history that could solve symbolic integration problems at the level of freshman calculus students, and was considered an AI project. Now that there are multiple systems on the market that can do much more than what SAINT did, many people do not consider these systems to be the results of AI research. The goal of machine intelligence differs from computational psychology and computational philosophy in being task-oriented rather than oriented toward the understanding of general intelligent behavior. A machine intelligence approach to a task is to use any technique that helps accomplish the task, even if the technique is not used by humans and would probably not be used by generally intelligent entities.

Subsymbolic AI

Computational psychology, computational philosophy, and machine intelligence are subareas of AI divided by their goals. AI researchers wander among two or all three of these areas throughout their career, and may even have a mixture of these goals at the same time. Cutting across these goals, however, is a recent division of approach into “symbolic” AI and “subsymbolic” AI. The key assumption of symbolic AI is that knowledge is represented by structures of semantically meaningful symbols. Each symbol representing some entity, be it abstract or concrete, that the intelligent system or agent is discussing, observing, reasoning about, or operating on. On the contrary, the key assumption of subsymbolic AI is that intelligent behavior can be attained without semantically meaningful symbols. Much of subsymbolic AI is included in the field of “soft computing”:

> “In contrast to the traditional, hard computing, soft computing is tolerant of imprecision, uncertainty and partial truth. The basic premises of soft computing are:
>
> - imprecision and uncertainty are pervasive
> - precision and certainty carry a cost
>
> The guiding principle of soft computing is:
>
> - exploit the tolerance for imprecision, uncertainty and partial truth to achieve tractability, robustness and low solution cost.
>
> Soft computing is not a single methodology; rather, it is a consortium or a partnership of methodologies. At this juncture, the principal constituents of soft computing (SC) are: fuzzy logic (FL), neurocomputing (NC) [*see* NEURAL NETWORKS], and genetic algorithms (GA) [*q.v.*]. The principal contribution of FL is a methodology for approximate reasoning and, in particular, for computing with words; that of NC is curve fitting, learning and system identification; and that of GA is systematized random search and optimization.” [Zadeh, 1995]

Heuristic Programming

Another way of distinguishing AI as a field is by noting the AI researcher's interest in *heuristics* (*q.v.*) rather than in *algorithms* (*q.v.*). Here I am taking a wide interpretation of a *heuristic* as any problem solving procedure that fails to be an algorithm, or that has not been shown to be an algorithm, for any reason. An interesting view of the tasks that AI researchers consider to be their own may be gained by considering those ways in which a procedure (*q.v.*) may fail to qualify as an algorithm.

By common definition, an algorithm for a general problem P is an unambiguous procedure that, for every particular instance of P, terminates and produces the correct answer. The most common reasons that a heuristic H fails to be an algorithm are that it doesn't terminate for some instances of P, it has not been proved correct for all instances of P because of some problem with H, or it has not been proved correct for all instances of P because P is not well-defined. Common examples of heuristic AI programs that don't terminate for all instances of the problem they have been designed for include searching (*q.v.*) and theorem proving (*q.v.*) programs. Any search procedure will run forever if given an infinite search space that contains no solution state. Gödel's Incompleteness Theorem states that there are formal theories that contain true but unprovable propositions. In actual practice, AI programs for these problems stop after some prespecified time, space, or work bound has been reached. They can then report only that they were unable to find a solution even though—in any given case—a little more work *might* have produced an answer. An example of an AI heuristic that has not been proved correct is any static evaluation function used in a program for playing computer chess (*q.v.*). The static evaluation function returns an estimate of the value of some state of the board. To be correct, it would return $+\infty$ if the state were a sure win for the side to move, $-\infty$ if it were a sure win for the opponent, and 0 if it were a forced draw. Moreover, for any state it is theoretically possible to find the correct answer algorithmically by doing a full minimax search of the game tree rooted in the state being examined. Such a full search is infeasable for most states, however, because of the size of the game tree. Nonetheless, static evaluation functions are still useful, even withoug being proved correct.

An example of a heuristic AI program that has not been proved correct because the problem for which it has been designed is not well-defined is any natural language understanding program or natural language interface. Since no one has any well-defined criteria for whether a person understands a given language, there cannot be any well-defined criteria for programs either.

Early History

Although the dream of creating intelligent artifacts has existed for many centuries, the field of artificial intelligence is considered to have had its birth at a conference held at Dartmouth College in the summer of 1956. The conference was organized by Marvin Minsky and John McCarthy, and McCarthy coined the name "Artificial Intelligence" for the proposal to obtain funding for the conference. Among the attendees were Herbert Simon (*q.v.*) and Allen Newell who had already implemented the Logic Theorist program at the Rand Corporation. These four people are considered the fathers of AI. Minsky and McCarthy founded the AI Laboratory at M.I.T.; Simon and Newell founded the AI laboratory at Carnegie-Mellon University. McCarthy later moved from M.I.T. to Stanford University, where he founded the AI laboratory there. These three universities, along with Edinburgh University, whose Department of Machine Intelligence was founded by Donald Michie, have remained the premier research universities in the field. The name Artificial Intelligence remained controversial for some years, even among people doing research in the area, but it eventually was accepted.

The first AI text was *Computers and Thought,* edited by Edward Feigenbaum and Julian Feldman, and published by McGraw-Hill in 1963. This is a collection of 21 papers, some of them short versions of Ph.D. dissertations, by early AI researchers. Most of the papers in this collection are still considered classics of AI, but of particular note is a reprint of Alan M. Turing's 1950 paper in which the Turing Test was introduced. (*see* TURING, ALAN.)

Regular AI conferences began in the mid to late 1960s. The Machine Intelligence Workshops series began in 1965 in Edinburgh. A conference at Case Western University in Spring, 1968 drew many of the U.S. AI researchers of the time, and the first biennial International Joint Conference on Artificial Intelligence was held in Washington, D. C. in May, 1969. *Artificial Intelligence,* still the premier journal of AI research, began publication in 1970.

For a more complete history of AI, see McCorduck 1979.

Neighboring Disciplines

Artificial Intelligence is generally considered to be a subfield of computer science, though there are some computer scientists who have only recently and grudgingly accepted this view. There are several disciplines outside computer science, however, that strongly impact AI and that, in turn, AI strongly impacts.

Cognitive psychology is the subfield of psychology that uses experimental methods to study human cognitive behavior. The goal of AI called computational psychology earlier is obviously closely related to cognitive psychology, differing mainly in the use of computational models rather than experments on human subjects. However, most AI researchers pay some attention to the results of cognitive psychology, and cognitive psychologists tend to pay attention to AI as suggesting possible cognitive procedures that they might look for in humans.

Cognitive science (*q.v.*) is an interdisciplinary field that studies human cognitive behavior under the hypothesis that cognition is (or can usefully be modeled as) computation. Although the overlap between AI and cognitive science is large, there are researchers in each field that would not consider themselves to be in the other. AI researchers whose primary goal is what was called machine intelligence earlier in this article generally do not consider themselves to be doing cognitive science, and cognitive science contains not only AI researchers, but also cognitive psychologists, linguists, philosophers, anthropologists, and others, all using the methodology of their own discipline on a common problem—that of understanding human cognitive behavior.

Computational linguists use computers, or at least the computational paradigm, to study and/or to process human languages. Like cognitive science, computational linguistics overlaps, but is not coextensive with, AI. It includes those areas of AI called natural language understanding, natural language generation, speech recognition and synthesis (*q.v.*), and machine translation (*see* LANGUAGE TRANSLATION), but also non-AI areas such as the use of statistical methods to find index keywords useful for retrieving a document.

AI-Complete Tasks

There are many subtopics in the field of AI—subtopics that vary from the consideration of a very particular, technical problem, to broad areas of research. Several of these broad areas can be considered *AI-complete,* in the sense that solving the problem of the area is equivalent to solving the entire AI problem—producing a generally intelligent computer program. Researchers in one of these areas may see themselves as attacking the entire AI problem from a particular direction. The following sections discuss some of the AI-complete areas.

Natural Language

The AI subarea of Natural Language is essentially the overlap of AI and computational linguistics (see above). The goal is to form a computational understanding of how people learn and use their native languages, and to produce a computer program that can use a human language at the same level of competence as a native human speaker. Virtually all human knowledge has been (or could be) encoded in human languages. Moreover, research in natural language understanding has shown that encyclopedic knowledge is required to understand natural language. Therefore, a complete natural language system will also be a complete intelligent system.

Problem Solving and Search

Problem solving is the area of AI that is concerned with finding or constructing the solution to a problem. That sounds like a very general area, and it is. The distinctive characteristic of the area is probably its approach of seeing tasks as problems to be solved, and of seeing problems as spaces of potential solutions that must be searched to find the true one or the best one. Thus, the AI area of search is very much connected to problem solving. Since any area investigated by AI researchers may be seen as consisting of problems to be solved, all of AI may be seen as involving problem solving and search.

Knowledge Representation and Reasoning

Knowledge representation (*q.v.*) is the area of AI concerned with the formal symbolic languages used to represent the knowledge (data) used by intelligent systems, and the data structures (*q.v.*) used to implement those formal languages. However, one cannot study static representation formalisms and know anything about how useful they are. Instead, one must study how they are helpful for their intended use. In most cases, this use is to use explicitly stored knowledge to produce additional explicit knowledge. This is what reasoning is. Together, knowledge representation and reasoning can be seen to be both necessary and sufficient for producing general intelligence—it is another AI-complete area. Although they are bound up with each other, knowledge representation and reasoning can be teased apart, according to whether the particular study is more about the representation language/data structure, or about the active process of drawing conclusions.

Learning

Learning is often cited as the criterial characteristic of intelligence, and it has always seemed like the easy way to produce intelligent systems: Why build an intelligent system when we could just build a learning system and send it to school? Learning includes all styles of learning, from rote learning to the design and analysis of experiments, and all subject areas. If the ultimate learning machine is ever created, it will acquire general intelligence, which is why learning is AI-complete.

Vision

Vision, or image understanding, has to do with interpreting visual images that fall on the human retina or the camera lens. (*see* COMPUTER VISION). The actual scene being viewed could be 2-dimensional, such as a printed page of text, or 3-dimensional, such as the world about us. If we take "interpreting" broadly enough, it is clear that general intelligence may be needed to do the interpretation, and that correct interpretation implies general intelligence, so this is another AI-complete area.

Robotics

The area of robotics (*q.v.*) is concerned with artifacts that can move about in the actual physical world and/or that can manipulate other objects in the world. Intelligent robots must be able to accommodate to new circumstances, and to do this, they need to be able to solve problems and to learn. Thus intelligent robotics is also an AI-complete area.

Integrated Systems

The most direct work on the problem of generally intelligent system is to use a robot that has vision and/or other senses, plans and solves problems, and communicates via natural language. Periodically throughout the history of AI, research groups have assembled such integrated robots as tests of the current state of AI.

Autonomous Agents

"Autonomous agents are computer systems that are capable of independent action in dynamic, unpredictable environments."[3] Autonomous agents are like integrated robots, except that they needn't have physical bodies nor need they

[3] Call for Papers, AGENTS '98, the Second International Conference on Autonomous Agents.

operate in the real, physical world. Instead, they may be completely software agents, sometimes called "softbots," and may operate in the world of the Internet, crawling around from file to file, collecting information for their human clients. To the extent that a softbot needs to be able to understand the files it comes across, it needs to be able to understand natural language, and to the extent that it needs to solve navigational problems to find the information it was asked for, it may need general purpose reasoning.

Applications

Throughout the existence of the field, AI research has produced spinoffs into other areas of computer science. Lately, however, programming techniques developed by AI researchers have found application to many programming problems. This has largely come about through the subarea of AI known as expert systems (*q.v.*). Whether or not any particular program should be considered intelligent or an expert according to the common use of those words is largely irrelevant to the workers in and the observers of the expert systems area. From their point of view they have tools and a methodology that are more useful for solving their problems than traditional programming tools and methodologies. From the point of view of AI as a whole, probably the best thing about this development is that after many years of being criticized as following an impossible dream by inappropriate and inadequate means, AI has been recognized by the general public as having applications to everyday problems.

References

1950. Turing, A.M. "Computing Machinery and Intelligence." *Mind,* **59** (October), 433–460.

1956. Newell, A. and Simon, H. A. "The Logic Theory Machine." *IRE Transactions on Information Theory,* **IT-2,** 61–79.

1963. Feigenbaum, E. A. and Feldman, J., Eds. *Computers and Thought.* New York: McGraw-Hill.

1963. Slagle, J. "A Heuristic Program that Solves Symbolic Integration Problems in Freshman Calculus." *Journal of the Association for Computing Machinery,* **10,** 507–520

1979. McCorduck, P. *Machines Who Think.* San Francisco: W. H. Freeman and Company.

1981. Barr, A. and Feigenbaum, E. A., Eds. *The Handbook of Artificial Intelligence,* Vol. I. Los Altos, CA: William Kaufmann, Inc.

1982. Barr, A. and Feigenbaum, E. A., Eds. *The Handbook of Artificial Intelligence,* Vol. II. Los Altos, CA: William Kaufmann, Inc.

1982. Cohen, P. R. and Feigenbaum, E. A., Eds. *The Handbook of Artificial Intelligence,* Vol. III. Los Altos, CA: William Kaufmann, Inc.

1989. Barr, A., Cohen, P. R. and Feigenbaum, E. A., Eds. *The Handbook of Artificial Intelligence,* Vol. IV. Reading, MA: Addison Wesley.

1990. Kurzweil, R. *The Age of Intelligent Machines.* Cambridge, MA: MIT Press.

1991. Shapiro, S. C., Ed. *Encyclopedia of Artificial Intelligence,* 2nd Edition. New York: John Wiley & Sons.

1995. Zadeh, L. "Foreword for Inaugural Issue." *Intelligent Automation and Soft Computing,* `http://www.laas.fr/autosoft/autosoft.html`.

1995. Russell, S. J. and Norvig, P. *Artificial Intelligence: A Modern Approach.* Englewood Cliffs, NJ: Prentice Hall.

1996. Doyle, J., Dean, T., *et al.* "Strategic Directions in Artificial Intelligence." *ACM Computing Surveys,* **28** (December), 653–670.

1997. Waltz, D. L. "Artificial Intelligence: Realizing the Ultimate Promise of Computing." In *Computing Research: A National Investment for Leadership in the 21st Century,* Washington, DC: Computing Research Association, 27–31.

Stuart C. Shapiro