1 Introduction

The opening section presents two very different types of attempt to provide a general characterization of AI. For Schank, AI is a distributed phenomenon: 'potentially . . . the algorithmic study of processes in every field of enquiry.' In the absence of a definition, he characterizes AI in terms of a list of features that he considers to be critical. He argues that the the bifurcation of AI into a scientific and an applications track is so decisive that 'the two routes have nothing to do with each other.' Finally, he lists and briefly discusses ten problem areas in AI that will not admit of solutions in the foreseeable future.

Chandrasekaran's paper lays out in historical perspective the methodological paradigms within which AI projects and explorations have at different times, and in different places, been pursued.

He takes the opening shots at connectionism and the echoes continue throughout this book culminating in the papers of section 9. He also introduces the 'symbolic' paradigm (another recurring theme) and attempts to clarify the issue of 'symbolic' and 'non-symbolic' representations. He offers the 'information processing level of abstraction' as a unifying paradigm.

He presents and discusses, with the aid of representative examples, three classes of theories in AI: architectural theories, logical abstraction theories, and general functional theories of intelligence. The paper concludes with a clear preference for the last class of theories and offers a specific functional-theory-type proposal about the nature of intelligence.

Within the discussion of the class of logic-based theories in AI, Chandrasekaran provides an overview and introduction to this important facet of the foundations of AI – the one that we deal with explicitly in the next chapter.

Cambridge University Press
978-0-521-35944-3 - The Foundations of Artificial Intelligence: A Sourcebook
Edited by Derek Partridge and Yorick Wilks
Excerpt
More information

What is AI, anyway?

Roger C. Schank

Artificial intelligence is a subject that, due to the massive, often quite unintelligible, publicity that it gets, is nearly completely misunderstood by people outside the field. Even AI's practitioners are somewhat confused with respect to what AI is really about.

Is AI mathematics? A great many AI researchers believe strongly that knowledge representations used in AI programs must conform to previously established formalisms and logics or else the field will be unprincipled and *ad hoc*. Many AI researchers believe that they *know* how the answer will turn out before they have figured out what exactly the questions are. They *know* that some mathematical formalism or other must be the best way to express the contents of the knowledge that people have. Thus, to them, AI is an exercise in the search for the proper formalisms to use in representing knowledge.

Is AI software engineering? A great many AI practitioners seem to think so. If you can put knowledge into a program, then that program must be an AI program. This conception of AI, derived as it is from much of the work going on in industry in expert systems, has served to confuse AI people tremendously about what the correct focus of AI ought to be, and about what the fundamental issues in AI are. If AI is just so much software engineering, if building an AI program means primarily the addition of domain knowledge such that a program "knows about" insurance or geology, for example, then what is to differentiate an AI program in insurance from any other computer program that works within the field of insurance? Under this conception, it is very difficult to determine where software engineering leaves off and where AI begins.

Is AI linguistics? A great many AI researchers seem to think that building grammars of English and putting those grammars on a machine is AI. Of course, linguists have never thought of their field as having very much to do with AI at all. But, as money for linguistics has begun to disappear, while money for AI has increased, it has become increasingly convenient to claim that work on language that had nothing to do with computers at all, has some computational relevance. Suddenly theories of language that were never considered by their creators to be process models at all, are now proposed as AI models.

Is AI psychology? Would building a complete model of human thought processes and putting it on a computer be considered a contribution to AI? Many AI researchers couldn't care less about the human mind. Yet, the human mind is the only kind of intelligence that we can reasonably hope to study. We have an existence proof. We know the human mind works. But, in adopting this view, one still has to worry about computer models that display intelligence but yet are clearly in no way related to how humans function. Are such models intelligent? Such issues inevitably force one to focus on the issue of the nature of intelligence apart from its particular physical embodiment.

In the end, the question of what AI is all about probably does not have only one answer. What AI is depends heavily upon the goals of the researchers involved. And any definition of AI is very dependent upon the methods that are being employed in building AI models. Last, of course, it is a question of results. These issues about what AI is exist precisely because AI has not yet been completed. They will disappear entirely when a machine begins to really be the way writers of science fiction have imagined it.

There are two main goals in AI that most practitioners would agree upon. First and foremost, the goal is to build an intelligent machine. And, second, the goal is to find out about the nature of intelligence. Both of these goals have at their heart a need to define intelligence. AI people are fond of talking about intelligent machines, but when it comes down to it, there is very little agreement on exactly what constitutes intelligence. And, it thus follows, there is very little agreement in AI about exactly what AI is and what it should be. We all agree that we would like to endow machines with an attribute that we really can't define. Needless to say, AI suffers from this lack of definition of its scope.

One way to attack this problem is to attempt to list some features that we would expect an intelligent entity to have. None of these features would define intelligence – indeed a being could lack any one of them and still be considered to be intelligent. Nevertheless each is an integral part of intelligence in its way.

Let me list the features I consider to be critical and then I shall briefly discuss them. They are: communication, internal knowledge, world knowledge, goal and plans, and creativity.

Communication: An intelligent entity can be communicated with. We can't talk to rocks or tell trees what we want, no matter how hard we try. With dogs and cats we cannot express many of our feelings, but we can let them know when we are angry. Communication is possible with them. If it is very difficult to communicate with someone, we might consider him to be unintelligent. If the communication lines are very narrow with a person, if he can only understand a few things, we might consider him to be unintelligent. No matter how smart your dog is, he can't understand when you discuss physics with him. This does not mean that he doesn't understand something about physics. You can't discuss physics with your pet rock either, but it doesn't understand physics at all. Your small child may know some physics, but discussions of that subject with him will have to be put in terms he can understand. In other words, the easier it is to communicate with an entity, the more intelligent the entity seems. Obviously there are many exceptions to this general feature of intelligence. There are people who are considered to be very intelligent who are impossible to talk to, for example. Nevertheless, this feature of intelligence is still significant, even if it is not absolutely essential.

Internal knowledge: We expect intelligent entities to have some knowledge about themselves. They should know when they need something; they should know what they think about something; and, they should know that they know it. Presently, probably only humans can do all this – we cannot really know what dogs know about what they know. We could program computers to seem as if they know what they know, but it would be hard to tell if they really did. To put this another way, we really cannot examine the insides of an intelligent entity in such a way as to establish what it actually knows. Our only choice is to ask and observe. If we get an answer that seems satisfying then we tend to believe that the entity we are examining has some degree of intelligence. Of course, this is another subjective criterion, a feature that when it is absent may signify nothing.

World knowledge: Intelligence also involves being aware of the outside world and being able to find and utilize the information that one has about the outside world. It also implies having a memory in which past experience is encoded and which can be used as a guide for processing new experiences. You cannot understand and operate in the outside world if you treat every experience as if it were brand new. Thus, intelligent entities must have an ability to see new experiences in terms of old ones. This implies an ability to retrieve old experiences which would have had to have been codified in such a way as make them available in a variety of different circumstances. Entities that do not have this ability can be momentarily intelligent but not globally intelligent. There are cases of people who are brain-damaged who perform adequately in a given moment, but forget what they have done soon after. The same is true of simple machines that can do a given job but do not know that they have done it, and have no ability to draw on that or other experiences to guide them in future jobs.

Goals and plans: Goal-driven behavior means knowing when one wants something and knowing a plan to get what one wants. There is usually a presumed correspondence between the complexity of the goals that an entity has and the sheer number of plans that an entity has available for accomplishing those goals. So, a tree has none or next to none of these, a dog has somewhat more, and a person has quite a few; very intelligent people probably have more. Of course, sheer number of recorded plans would probably not be a terrific measure of intelligence. If it were, machines could easily be constructed that met that criterion. The real criterion with respect to plans has to do with interrelatedness of plans and their storage in an abstract enough way as to allow a plan constructed for situation A to be adapted and used in situation B.

Creativity: Finally, every intelligent entity is assumed to have some degree of creativity. Creativity can be defined very weakly, including for example, the ability to find a new route to one's food source when the old one is blocked. But, of course, creativity can also mean finding a new way to look at something that changes one's world in some significant way. And, it certainly means being able to adapt to changes in one's environment and being able to learn from experience. Thus, an entity that doesn't learn is probably not intelligent, except momentarily.

Now, as I said, one needn't have all of these things to be intelligent, but each is an important part of intelligence. That having been said, where do current AI programs fit in? It seems clear that no AI model is very creative as of yet, although various ideas have been proposed in this regard lately. It also seems clear that no AI models have a great deal of internal knowledge. In general, AI programs don't know what they know, nor are they aware of what they can do. They may be able to summarize a news wire, but they don't know that they are summarizing it.

On the other hand, programs that have goals and plans to accomplish those goals have been around since the inception of AI. Work on such programs has spawned a variety of ideas on how planning can be accomplished, particularly within the domain of problem solving. Programs that have external knowledge usually have not been considered to be part of AI at all. Database retrieval is not in any way connected with AI, although it has seemed clear to AI researchers that they must eventually concern themselves with how knowledge is best organized in order to have really intelligent machines. Nevertheless, many programs for organizing and retrieving knowledge do, of course, exist.

Programs that communicate with computers have been around as long as there have been computers, of course. But this communication has been less than satisfactory. Most non-computer professionals complain bitterly about the difficulty in getting a computer to do what is wanted, and of course, the computer industry has been responsive to this, producing better and better interfaces. But, in the end, computers will not really be easy to use until they can see, hear, read, and generally understand what we say to them and what we want them to do.

In AI, these subjects have always been considered to be important parts of the field and much research has been done on them.

As AI has become more commercialized, the parts of AI research that have been the most advanced in terms of engineering would, one might have imagined, become those areas where the commercial action would begin. But, as often happens, salesmanship and market readiness often determine what gets sold. So AI entered the world through the creation of so-called expert systems, which were engineering attempts to take some of the problem solving and planning models that had been proposed in AI and give them real world relevance. The problem was that these experts lacked what I have termed internal knowledge and creativity. And, it is very difficult to have an expert who doesn't know what he knows, how he came to know it, or how to adapt if circumstances are somewhat different than they were supposed to be. Most of all, experts with no memories are no experts at all.

Partly as a result of the commercialization of expert systems, equating AI with expert systems in the public eye, and partly as a result of the usual battles AI has always faced with older fields of inquiry that relate to it, AI is in a serious state of disruption.

Most AI people seem to have chosen one of two routes, to get them out of their state of confusion. The first of these routes I will call the *applications route*. In this view of AI, the job is to build real working systems. Whether these systems are AI or not loses its import as one begins to work on them. The problem is make them work at all, not to be a purist about what is or is not AI. As anyone who has ever worked on a large software engineering program knows, this task is so complex as to make all other problems pale by comparison. Making big programs work is hard. And when they are finished are they AI? Does it matter?

The second route is what I will call the *scientific route*. This route sounds good in principle and it has as its premise a desire to avoid the commercialization of AI and work only on impossible problems like the brain, or neat problems like logic. Let the applications route people do as they will, the scientific route people have chosen simply to ignore them and bolt the door.

Thus, without actually deciding to do so, AI has made a decision. Either one defines AI as a modern methodological tool now being used in the ancient enterprise of the study of mind, the *scientific answer*, or, one's definition of AI is, in essence, the *applications answer*, namely an attempt to create certain new computer technology that relates to some behaviors previously done only by humans.

This seems fine in principle – many fields have a scientific, theoretical group and an applications group that derives its work from the scientific work. And, this would be nice in AI too, if it were the case. What actually is the case is that the scientific workers are, for the most part, concerned with issues that are very far away from potential applications, and the applications folk have been busy applying results from earlier days which are known to be seriously inadequate. This does not mean that they are not building useful applications, sometimes they are. But, it does mean that, for all intents and purposes, the two routes have nothing to do with each other.

One problem with the applications answer is that it is very imprecise. Is all new computer technology to be labeled AI? Certainly, if one reads the advertisements in the computer magazines, it is easy to believe that AI is anything anyone says it is, that there is no definition. But, to an AI researcher (as opposed to an AI businessman), only a small fraction of advances in computer software and hardware would seem to qualify as advances in AI. The technology that AI people want to create usually involves solving some fundamental problem, and the solution itself involves decisions on the nature of what kinds of things are part of a computer program. Further, AI usually means getting a machine to do what previously only humans have done before (rather than simply improving existing techniques). The problem with this definition has been obvious to AI people for some time. As soon as something radically new has been accomplished, then, since computers have at that point done it, it is thus no longer uniquely human, and thus no longer AI. So, one question that needs to be answered on the technological side is, "Can some definition as to the nature of AI software be made such that, under all circumstances, it will be seen as uniquely part of or derivative from AI?"

What is really the case is that it is not possible to define very clearly which pieces of new software are AI and which are not. In actuality, AI must have an issues-related definition. In other words, people do arithmetic and so do computers. The fact is, however, that no one considers a program that calculates to be an AI program, nor would they, even if that program calculated in exactly the way that people do. The reason that this is so is that calculation is not seen as a fundamental problem of intelligent behavior and also that computers are already better at calculation than people are. This two-sided definition, based upon the perception of the fundamental centrality of an issue with respect to its role in human intelligence, and the practical viewpoint of how good current computers are at accomplishing such a task already, constitutes how one defines whether a given problem is legitimately an AI problem. For this reason, much of the good work in AI has been just answering the question of what the issues are.

Or, to put this another way, what is AI is defined not by the methodologies used in AI, but by the problems attacked by those methodologies. A program is not an AI program because it uses LISP or PROLOG certainly. By the same token, a program is not an AI program because it uses some form of logic or if-then rules. Expert systems are only AI programs if they attack some AI issue. A rule-based system is not an AI program just because it uses rules or was written with an expert system shell. It is an AI program if it addresses an AI issue.

One thing about AI issues though, is that they change. What was an issue yesterday may not be one today. Similarly, the issues that I believe to be critical today may disappear ten years from now. Given that that is the case, defining AI by issues can make AI a rather odd field, with a constantly changing definition. But, there are some problems that will endure, that tend to define AI. I will discuss some of these below:

1 Representation
2 Decoding
3 Inference
4 Control of combinatorial explosion
5 Indexing
6 Prediction and recovery
7 Dynamic modification
8 Generalization
9 Curiosity
10 Creativity

1. **Representation** Probably the most significant issue in AI is the old problem of the representation of knowledge. *What do we know?*, and *how do we get a machine to know it?* is the central issue in AI. An AI program or theory that makes a statement about how knowledge that is of a generality greater than the range of knowledge covered by the program itself ought to be represented is a contribution to AI.

2. **Decoding** It is of no use to have a very nice knowledge representation if there is no way to translate from the real world into that representation. In natural language, or vision systems, for example, decoding is often the central problem in constructing an AI program. Sometimes, of course, the decoding work is so difficult that the programmers forget to concern themselves with what they are decoding into, that is, what the ideal representation ought to be, so that they make the work harder for themselves. Deciding that the representation is a given fact, that it is predicate calculus, or syntactic phrase markers, for example, can complicate the problem, relegating the decoding work to some other, often non-existent, program.

3. **Inference** Information is usually more than the sum of its parts. Once we have decoded a message (visual, verbal, symbolic, or whatever) we must begin to attempt to extract the content of that message. Usually the content is much more than has been expressed directly. We don't say every nuance of what we mean. We expect our hearer to be smart enough to figure some of it out for himself. Similarly, we must attempt to figure out the significance of what we have seen, making assumptions about what it all means. This is the problem of inference.

Human memory is highly inferential, even about prior experiences, and retrieval of information. People are capable of answering questions from very incomplete data. They can figure out if they should know something and whether they might be able to figure it out. Such self-awareness depends strongly upon an ability to know how the world works in general, or, the representation problem again. Building a program that knows if it would know a thing is a very important task.

4. **Control of combinatorial explosion** Once you allow a program to make assumptions about what may be true beyond what it has been told, the possibility that it could go on forever doing this, becomes quite real. At what point do you turn off your mind and decide that you have thought enough about a problem? Arbitrary limits are just that, arbitrary. It seems a safe assumption that it is the structure of our knowledge that guides the inference process. Knowing what particular knowledge structure we are in while processing can help us to determine how much we want to know about a given event. Or, to put this another way, contexts help narrow the inference process. There are many possible ways to control the combinatorics of the inference process – deciding among them and implementing them is a serious AI problem if the combinatorial explosion was started by an AI process in the first place.

5. **Indexing** It is all well and good to know a great deal, but the more you know, the harder it should get to find what you know. The most knowledgeable man on earth should also be the slowest to say anything, by that reasoning. This is called the paradox of the expert in psychology. It is a paradox precisely because it is untrue. Obviously, people must have ways of organizing their knowledge so that they can find what they need when they need it. Originally this problem was called the search problem in AI. But, viewed as a search problem, the implication was that faster search methods were what was needed. This would imply that experts were people who searched their data bases quickly and that seems quite absurd. It is the organization and labeling of memory and episodes in memory that is the key issue here. For any massive system, that is for any real AI system, indexing is a central, and possibly the central problem. AI programs are not usually large enough to make their answers to the indexing question meaningful, but the construction of programs of the appropriate size should become more important in the years ahead.

6. **Prediction and recovery** Any serious AI program should be able to make predictions about how events in its domain will turn out. This is what understanding really means, that is, knowing to some extent what is coming. When these predictions fail, which they certainly must in any realistic system, an intelligent program should not only recover from the failure, but it must also explain the failure. That is, programs must understand their own workings well enough to know what an error looks like, and be able to correct the rule that caused that error in addition to being able to recognize that situation when it occurs again. As an example

of the kind of thing I am talking about, a computer should be able, by use of the same basic scientific theory, to do an adequate job of forecasting stocks or weather, or playing a game of chess, or coaching a football team. What I mean by *the same basic theory* is that the theory of prediction, recovery from error, error explanation, and new theory creation should be identical in principle, regardless of domain.

7. **Dynamic modification** AI went through a long period of trying to find out how to represent knowledge. We needed to find out what was learned before we could even consider working on learning itself. But, most of us have always wanted to work on learning. Learning is, after all, the quintessential AI issue. What makes people interesting, what makes them intelligent, is that they learn. People change with experience. The trouble with almost all the programs that we have written is that they are not modified by their experiences. No matter how sophisticated a story understander may seem, it loses all credibility as an intelligent system when it reads the same story three times in a row and it fails to get mad, bored, or even to notice. Programs must change as a result of their experiences or else they will not do anything very interesting.

Similarly, any knowledge structures, or representations of knowledge that AI researchers create, no matter how adequately formulated initially, must change over time. Understanding how they are changed by actual use during the course of processing information is one of the major problems in representation itself. Deciding when to create a new structure or abandon an old one is a formidable problem. Thus, new AI programs should be called upon to assimilate information and change the nature of the program in the course of that assimilation. Clearly such programs are necessary before the knowledge acquisition problem can be adequately attacked. It should also be clear that an AI program that cannot build itself up gradually, without requiring all its knowledge stuffed in at the beginning, is not really intelligent.

I will now give a definition of AI that most of our programs will fail. AI is the science of endowing programs with the ability to change themselves for the better as a result of their own experiences. The technology of AI is derived from the science of AI and is, at least for now, unlikely to be very intelligent. But, it should be the aim of every current AI *researcher* to endow his programs with that kind of *dynamic* intelligence.

8. **Generalization** A program that can form a generalization from experience that can be tested would be of great significance. This program would have to be able to draw conclusions from disparate data. The key aspect of a good generalization maker is his ability to connect together experiences that are not obviously connectable. This is the essence of creativity. A key AI problem, therefore is to understand new events and make predictions about future events by generalizing from prior events. These generalizations