AI Magazine Volume 4 Number 3 (1983) (© AAAI)

Artificial Intelligence:

An Assessment of the State-of-the-Art and Recommendation for Future Directions

David Waltz, Chairman[1]

NSF *Information Technology Workshop*
Xerox International Center
Leesburg, Virginia
January 5-7, 1983

Abstract

This report covers two main AI areas: natural language processing and expert systems The discussion of each area includes an assessment of the state-of-the-art, an enumeration of problems areas and opportunities, recommendations for the next 5-10 years, and an assessment of the resources required to carry them out A discussion of possible university-industry-government cooperative efforts is also included

ARTIFICIAL INTELLIGENCE (AI) is the attempt to (1) understand the nature of intelligence and (2) produce new classes of intelligence machines through programming computers to perform tasks which require reasoning and perception. The goal of AI as a whole is to produce machines that act intelligently. By "act intelligently," we mean to cover a broad range of activities, only some of which are directly human-like; our machines may eventually be far better at certain intelligence tasks than people are (much as a calculator does long division better than people do), yet may lack other human characteristics (e.g., ambition, fear, general world knowledge, mobility). Important characteristics that such machines would have to have before AI could be said to have succeeded include common sense; the ability to learn from experience; the ability to accept, generate and act appropriately on natural language input; perception and general situation assessment; and so on.

[1]Authors of the report are as follows: David Waltz, Chairman, University of Illinois, Urbana; Michael Genesereth, Stanford University; Peter Hart, Fairchild Advanced Research and Development Laboratory; Gary Hendrix, Symantec, Inc ; Aravind Joshi, University of Pennsylvania; John McDermott, Carnegie-Mellon University; Thomas Mitchell, Rutgers University; Nils Nilsson, SRI International; Robert Wilensky, University of California, Berkeley; and William Woods, Bolt Beranek & Newman, Inc

Coverage of this Report

For the purposes of this report, we concentrate on natural language (NL) processing and expert systems.[2] Both of these areas have near term practical potential, and yet are

[2] Because we have researchers primarily from these two areas, in turn because these two areas were specifically selected for concentration by the workshop steering committee However, there are many other important problems in AI that are not well-represented here The areas of computer perception and pattern analysis are particularly unfortunate omissions Other areas of major AI research are included to some degree, because they overlap other sections of this report (robotics, parallel and distributed hardware, software and programming environments, and information sciences). In each of these areas, however, AI has specific needs, often somewhat different from the rest of computer science, which deserve consideration We hope that these important AI research areas will be considered in future assessment and planning meetings, especially because progress in natural language and expert systems ultimately depends critically on virtually all other AI areas

sufficiently open-ended to remain interesting to researchers past the next ten years

The biggest AI news of the recent past has been the commercial introduction and industrial use of the number of AI systems, especially NL and expert systems. This news is significant because (1) it has quieted critics who argued that AI would never produce useful results, and (2) the applications themselves have high intrinsic value. Some specific systems include:

INTELLECT (AI Corp., Waltham, MA) A natural language system that can be added to a customer's existing data base.

STRAIGHT TALK (Symantec for Dictaphone Corp.) A natural language data base system embedded in a word processor, intended for uses such as the storing of address, phone number, organization, and salary information

R1 (Designed at Carnegie-Mellon University for Digital Equipment Corp.) An expert system that can check and correct the configuration of the 450 or so components that can go into a VAX computer

PROSPECTOR (SRI International) An expert geologist system that can find commercially exploitable mineral deposits from assay data

The Machine Intelligence Corp. Vision module (based on research done at SRI), a system that can be "taught" by a non-programmer to recognize and act upon a variety of parts on an assembly line, when used with a robot arm

DENDRAL (Stanford) An expert system fo discovering the molecular structure of organic compounds from mass spectrogram data.

MACSYMA (MIT) A general mathematical aids system, including the abilities to do symbolic integration, factoring, simplification, plotting, and much more.

MOLGEN (Stanford) An expert system for designing experiments in molecular biology

In addition, there are a number of systems that, while not being marketed, have received fairly wide use, for instance:

MYCIN (Stanford) An expert system for medical diagnosis and treatment prescription

EMYCIN (Stanford) MYCIN with specific medical knowledge removed, EMYCIN has been used by non-programmers to make expert systems in other areas such as tax advising and estate planning

INDUCE/PLANT (University of Illinois) A system that has learned to generate its own rules for diagnosing soybean diseases from examples presented to it with better performance than most experts

Dipmeter advisor (Schlumberger-Doll) A system for interpreting oil-well log data

LISP machines (MIT; Symbolics, Inc; LMI, INC.; Xerox; BBN) Computers specialized for the LISP language that pioneered the ideas of the personal workstation and user-friendly programming environments

There are in addition a number of companies that plan to introduce or use internally AI products in the near future, including NL systems from Texas Instruments, Cognitive Systems Inc., Symantec, and Hewlett-Packard, and expert systems from Teknowledge, Fairchild, Intelligenetics, Schlumberger, and others.

Overview of this report

In the sections that follow, we first discuss in some detail the nature of the most important current technical and scientific issues in natural language processing and in expert systems. We then discuss organizational problems, including difficulties involving manpower, equipment, funding, education, and cooperation between industry, government, and universities. This is followed by a list of recommendations that we believe can aid in keeping the U.S. lead in these important areas.

Natural Language Research

Endowing computers with an ability to communicate with humans in natural language (e.g. ordinary English) has been a major topic of research in AI for over 20 years The ultimate goal of creating machines that can interact in a facile manner with humans remains far of, awaiting breakthroughs in basic research, improved information processing algorithms, and perhaps alternative computer architectures. However, the significant progress experienced in the last decade demonstrates the feasibility of dealing with natural language in restricted contexts, employing today's computers.

Continuing research in this area seems likely to lead both to progressively more practical, cost-effective systems and to a deeper understanding of the natural phenomena of NL communication. Each of these goals has importance in isolation; pursuing them simultaneously enables progress on each to support progress towards the other.

Application Areas

Natural language processing has a broad range of possible application areas.

Machine Translation [MT] MT involves using machines to convert documents written in one natural language to corresponding documents written in another language, but with equivalent meaning. MT was proposed in 1946 and became a forerunner of today's work in AI Creation of fluent translations remains an elusive goal.

Document understanding Document understanding involves reading documents by machine, and assimilating the information the documents contain into a larger framework of knowledge After reading a document a device of this sort might produce an abstract of it, alert people who are likely to be interested in it, or answer specific questions based on the information it contains If such a device has read and assimilated many documents, it might act as a librarian, directing users to especially pertinent references

Document preparation aids. These aids could perform the task of an experienced editor, detecting errors in spelling and grammar, and suggesting ways to rephrase passages of text to make them more understandable and to make them conform to the patterns of high quality language usage

Document generation This task, related to document understanding, involves translating information stored in a formal language in a computer's memory into ordinary language The documents produced might be single sentences or extensive texts For example, information encoded in a formal language regarding the repair of an electromechanical device could be used as the basis for mechanically generating instruction manuals in a variety of natural languages Moreover, from the same formal description, different manuals could be generated for different audiences such as end users, repair personnel, and engineers Ultimately, documents could be tailored to the background of each particular individual, making each document more understandable and generating the correct level of detail.

Systems control This is the applications area with the greatest promise for near-term achievement It involves the use of NL in the control of computer systems. By coupling a natural language interface with different types of devices, a range of possible systems may be produced, including systems that (1) provide answers to questions by accessing large data bases; (2) control such complex systems as industrial robots, power generators, or missile systems; (3) furnish expert advice about medical problems, mechanical repairs, mineral exploration, the design of genetic experiments, or investment analysis; (4) create graphical displays; (5) teach courses in a broad range of subjects, interacting with students in English

Practical systems for (1) above are already commercially available, and rapid development of this area can be expected over the next several years. Little work has been done on area (2), but there appears to be no special technical obstacles to producing elementary systems in this area if programs are undertaken Area (3) involves integrating work on natural language processing with work on expert systems; some limited demonstration programs using stylized input and canned output have been produced already, but much remains to be done. Areas (4) and (5) require considerable new work, but significant progress could be made if a concentrated effort were undertaken. (Note: Work in this area could help educate people in computer science in general and in AI in particular)

In considering natural language as the command language for controlling computer systems, it is important to keep in mind that English is well-suited to some kinds of man-machine interaction and poorly suited for others (such as those involving extensive manipulation of numbers) English is often useful:

- when dealing with computer-naive users,
- for tasks which computer experts do infrequently,
- in situations where English is more concise than formal language, and
- for activities in which natural language is the subject of analysis, e g intelligence gathering

Speech understanding Speech understanding involves the coupling of natural language processing with acoustic and phonetic processing to achieve a device that can understand spoken as opposed to typed input. Advances in this area are currently inhibited primarily by the lack of satisfactory acoustic/phonetic devices for recognizing sequences of individual words in continuous speech. Special purpose parallel hardware for word recognition is being pursued by reseachers outside AI

Importance and Economic Impact

Perhaps the most important economic factor of the current age is that, as a society, we are moving away from an economy based on the manufacture and dissemination of goods to an economy based on the generation and dissemination of information and knowledge. Much of the information and knowledge is expressible in English and much of the task of gathering, manipulating, multiplying, and disseminating it can be greatly aided by computers. Thus research in NL understanding can have a two-fold positive impact in our shifting economy:

1 NL can enable computers to interact with users in ordinary language, and therefore it can make computer power available to segments of the population that are unable or unwilling to learn one of the formal languages usually required for interaction with computers

2 NL can increase knowledge productivity in providing mechanical means for manipulating knowledge expressed as natural language text.

State-of-the-Art in Natural Language Processing

Currently, we understand how to do a reasonably good job of literal interpretation of English sentences in static contexts and limited, well-structured domains of application This is not to say that there are no open problems in this area, but rather, compared with the other aspects of language to be discussed, there is a substantial body of known results and proven techniques.

A number of good application areas are now possible. Examples include NL data base front ends; NL interfaces for expert systems, operating systems, system HELP facilities, library search systems, and other software packages; text filters, text summarizers; machine-aided translation; and grammar checkers and critics

Parsing algorithms for syntactic analysis of sentences and techniques for semantic interpretation (to determine literal meaning) are well developed, making possible many practical applications

It is now possible to think about a much wider range of types of NL processing because machines have become substantially larger and address spaces have become reasonable for NL systems In the past, a great deal of effort was devoted to attempts to collapse code length so that space could be made available to expand the capabilities of NL programs This is no longer such a problem, and we can trade off space for ease of programming.

However, extensive resources are required to develop a new NL application even in areas we understand quite well, because there is no such thing as a small natural language system. If a system is to accept *natural*language, that is, unrestricted text or what people naturally think of saying in the manner they think of saying it, it must have a large vocabulary, a wide range of linguistic constructions, and a wide range of meaning representations A small system cannot be very natural

Existing techniques begin to break down when we begin to scale up to more open-ended applications where inherent limitations of the domain are no longer adequate to resolve ambiguities in the language, or where sophisticated discussion of time or three-dimensional space is required, or where discussion and modelling of the beliefs, goals and rational behavior of intelligent agents is required

Research Areas in Natural Language

The most active current areas of research (and the most promising for new breakthroughs) lie in the area of recognizing the intent of speakers and the relationships between sentences in continuous discourse, taking into account the structure of the preceding discourse, the non-linguistic aspects of the situation of utterance, and models of the beliefs and goals of the communication agents.

Historically our understanding of the phenomena of language has moved gradually from the most visible and salient aspects such a phonology, lexicon, and syntax towards the more internal and intellectual elements of semantics, pragmatics, and reasoning Like the stages of Piagetian development it has seemed that we must first gain a mastery of the earlier stages before proceeding to explore the later ones.

Natural Language Planning. A recent innovation, with enormous potential for increasing our understanding and capabilities in NL has been the evolution of a methodology in which communication is treated as a special case of a goal-oriented action in a common framework with nonlinguistic actions This allows for planning and reasoning about goal-oriented activities which involve communication and acquisition of information as a part of the overall activity of a system. As a simple example, in order to get into a locked room occupied by a human, a robot might construct and execute a plan that would lead to its saying, "Please unlock the door so I can come in."

Speaker's Intent and Plan and Recognition The planning approach to the problem of communication has put new flesh on the basic skeleton of the theory of speech acts advanced by philosophers Grice, Austin, and Searle and promises significant advances in linguistic fluency of communicating machines It also provides a framework for nonlinguistic communication through actions such as pointing or displaying a picture There are, however, substantial technical problems that must be addressed in order to realize the promise of this approach. These include development of reasoning systems capable of modelling and reasoning about the beliefs, goals, and actions of rational agents, the representation, organization and access of the knowledge necessary to support such reasoning, and the discovery of frameworks, methods and algorithms capable of combining syntactic, semantic, pragmatic and general world knowledge to perform the overall task of understanding an utterance of context.

Representation and Commonsense Reasoning. The emerging focus on problems of interpretation in context have suddenly put great stress on the problems of knowledge representation and commonsense reasoning The solution of these problems requires extensive use of general world knowledge and knowledge about the rational behavior of intelligent, communicating agents. It is impossible to overemphasize the importance of representation and reasoning for this new (and final?) stage of learning about language.[3] To extend existing systems we must address many fundamental problems

1. We must decide exactly what knowledge to include in an NL system. This type of knowledge is considered by people to be "commonsense," and is not codified anywhere
2. We must commit ourselves to "primitive" elements, that is, items that can be used in definitions but that will not themselves be defined (except perhaps by reference to sensory systems)
3. We must devise appropriate notational systems
4. Knowledge must be organized in memory to facilitate access In particular, facts may need to be recalled upon demand, and facts relevant to the current context may need to be inferred automatically
5. Knowledge and reasoning about the physical world is required for determining the referents of noun phrases, interpreting prepositional phrases, disambiguating word senses and generally understanding

---

[3]The research areas of knowledge representation and commonsense reasoning are areas likely to have considerable impact on expert systems work as well Current expert systems are based on representations chosen specifically for an application, with little general scientific understanding of the power or limitations of those representations or the consequences which may emerge at later stages of development from representational decisions made at the outset

text that refers to some physical situations This kind of knowledge is also critically important for judging whether NL is literally plausible, as opposed to metaphorical, humorous, sarcastic, or in error.

Generation. The areas described so far have treated language as an input modality only, i e they have concentrated on language understanding. Current NL systems do not often use NL as an output modality, but this situation is expected to change quite radically in the future. Systems of the future will have to deal with situations where the system understands that a user doubts, is confused by, or does not understand the system's response. Such systems must be prepared to paraphrase or explain their responses. There is already a recognized need for all expert systems to be able to explain their reasoning and conclusions (see below)

There are three main aspects of generation: (1) deciding what to say, (2) deciding how to say it, and (3) finally saying it. Thus it should be clear that the previously mentioned research areas (especially planning, knowledge representation, and commonsense reasoning) are highly relevant to generation as well as to understanding.

Algorithms. The study of the formal properties of algorithms needed to manipulate the various types of NL representations mentioned above is itself an important research activity. We know something about parsing algorithms and algorithms for certain logic systems We will need to know much more about algorithms, for example, those that allow us to combine syntactic, semantic, and pragmatic knowledge One recent example is the discovery that grammatical representations that combine syntax and semantics (that is, provide grammatical representations that are convenient for semantic interpretation) have properties which make their parsing only slightly more complex that the parsing of context-free grammars

Scaling up – Learning. If a NLP (Natural Language Processing) system is to be useful, it must be able to handle a large vocabulary and have access to knowledge base. This imposes two constraints upon the design of an NL system First, it must operate efficiently when it is in possession of relatively large amounts of knowledge. Second, means must be found for building up large knowledge bases appropriate for NL understanding, and for progressively expanding a system.

Most working AI systems are of limited scope, and so there has been little actual experience with truly large scale AI systems Therefore, it is an act of faith to assume that our current techniques will be effective in vastly larger systems. One possible avenue for research involves the further investigation of techniques from data base management, in which problems of scale are routine, but in which the data is much less complexly structured. Techniques specific to the management of large collections of complex knowledge need to be developed.

The desirability of a large and incrementally expandable system poses some quite general questions of system design In addition, some problems arise that are singular to NLP. For example, it would be desirable to build a system that could learn both vocabulary and world knowledge be dialog in NL with a user, or by reading text (e.g dictionaries, encyclopedias, texts, stories, etc.). The simpler aspects of this problem are within our current understanding. However, at the extreme end, one encounters problems of machine learning that require fundamental advances in our basic understanding.

Opportunities in Natural Language

Having reviewed the state-of-the-art and the major research problems for the next ten years, we would like in this section to briefly state some of the short term goals that could lead to significant improvements in NL technology The highest priority should be given to the basic research issues enumerated in the section above, so there will be new application areas ten years.

Although the research areas described above will require ten years or longer to arrive at [general] solutions, it is possible to make incremental contributions toward the limited handling of various discourse phenomena, such as ellipses (i.e. the omission of words that can be understood in context), recognizing a user's intent in restricted situations, discovering some of the user's beliefs from the presuppositions of the user's input, modelling aspects of commonsense, learning by reading text or engaging in dialogue, etc. It should be emphasized, however, that one should not expect progress in the basic research topics to be strictly incremental. There will have to be some major breakthroughs in order to obtain general and principled solutions for these topics.

Research and development of good NL tools should pay excellent dividends We now have available a few parsers and knowledge representation systems, which should make it possible to build new systems more rapidly, by using off-the-shelf components for programs. We do need to have these systems well-documented, and we still need more experience in tailoring systems from these components, but there is the opportunity to begin building custom NL systems now, at last for limited task domains

Because it is now possible to relatively easily produce special-purpose chips, the identification of potential parallelism in NL processing has gained importance. We can realistically consider algorithms for highly parallel words sense selection, truly concurrent syntactic, semantic and pragmatic evaluation of sentences, speech format extraction, etc with the expectation that such work can lead to novel machine architectures more appropriate for NL processing

Many good novel applications are possible within the next ten years, provided that we can solve some of the basic research problems enumerated above Many of these arise in conjunction with "information utilities" (i.e information services available via phone or cable connections) Possible public services include automatic directories of names, addresses, yellow pages, etc.; electronic mail; on-line catalogues and ordering facilities; banking and tax services; routing directions; access to books and periodicals, through titles, authors, topics, or contents; and undoubtedly many others. All of these services could also have on-line NL help facilities and manuals. There are parallel needs in business and in the military services, e.g., for command and control, inventory control, ordering and shipping, coordination of organization planning and plan execution, and so on

Other good application bets concern the control of systems via NL. For example, we should, within ten years, be able to solve the problem of instructing robots in NL, much as one would instruct a human assistant. This can allow the robot to understand the goals of the instruction, rather than just the means for achieving the goals, as is the case now with teaching-by-leading-through-motions, or "teaching" by ordinary programming Understanding goals would help a robot in dealing with a wider range of situations and in recovering from errors.

Research on speech understanding should have a high priority. Once continuous speech understanding is possible, the number of good NL application areas will grow dramatically. Provided that sufficient funds are made available, continous speech systems seem close enough to reality that we should begin to think more seriously now about the realm of new application areas that speech systems will open up. It is critically important that we at the same time make substantial progress on the fundamental problems in NL understanding, so that we can move rapidly to produce useful systems Speech recognition alone, without NL understanding and action components, would be of little value

Expert Systems

Expert systems research is concerned with the construction of high performance programs in complex domains. By "high performance" we mean functionality and efficiency comparable with or better than the best human experts. By "complex domains" we mean those application areas requiring substantial bodies of knowledge, often of an uncertain or judgmental nature. This potential for capturing judgmental knowledge expands the range of problems to which computers have been and potentially could be applied.

What differentiates the expert system methodology from traditional computer programming is an emphasis on the symbolic manipulation capabilities of computers, in particular the declarative representation of world knowledge, the explicit encoding of heuristics, and the exploration of non-numeric data structures for computer simulations

State-of-the-Art

As just mentioned, expert systems differ from more conventional computer programs (which, like econometric models, for example, may possess considerable "knowledge") by their ability to deal with uncertain and judgmental knowledge, represented in a symbolic (often declarative) form Expert systems differ from other AI programs, such as natural language understanding programs, because of their concern for subjects that usually require specialized training in a professional field, such as medicine, law, mathematics, or computer circuit design. It is perhaps one of the most surprising developments of the past ten years that a useful portion of the expert knowledge required for high-level performance in many of these fields can in fact be encoded in computer programs.

In addition to the high performance, most of these systems can explain their conclusions so that their users have a better understanding on which to base action and greater confidence in the quality of the systems' decisions This feature can best be explained by considering two contrasting approaches that could be used in constructing a program for medical diagnosis. In a system based on statistical decision theory using joint frequency distributions over symptoms and diseases, for example, the physician user might be informed only that, given the symptoms, the most likely disease is diabetes A physician would have to take the conclusions on faith, and even if it were highly accurate, a user would be reluctant to depend blindly on such a system. This use of unexplained conclusion is in sharp contrast to what most expert systems can provide Current systems can describe the "chain of reasoning" employed by the system in reaching its conclusion This chain refers to the judgments, assumptions, rules and intermediate conclusions used by the system. Physicians find these kinds of explanations absolutely essential when deciding whether to rely on the machine's diagnosis or not

Despite their success, current expert systems suffer from a variety of limitations. Among those shortcomings that ought to yield to concentrated research efforts during the next ten years or so are the following: overly narrow domains of expertise; inadequate communication channels with the user (e g. need for better natural language and graphics); inability to represent certain kinds of knowledge easily (e.g. knowledge about processes, time, three-dimensional space, beliefs of the user); and th great difficulty of building and modifying the expert knowledge bases on which these systems are based.

It is this last area, knowledge acquisition, where we can expect the most difficulty. Current systems are built by having computer scientists interview experts in the domain of application. The knowledge obtained from these experts, usually in the form of English sentences, must then be structured by the computer scientist (often called a "knowledge engineer") so that it can be unambiguously and economically represented in the computer. (Incidentally, this process of precisely structuring knowledge, for instance geological knowledge, for the computer can just as well be thought of as an advance in the science of geology as an exercise in system engineering) Although it is reasonable to expect that we will be able to develop sophisticated computer aids for the knowledge acquisition process, it must be remembered that to completely automate the knowledge acquisition process will require rather dramatic advances in other areas of AI such as natural language understanding and generation, and machine learning.

Other problems that will require many years of work involve connecting expert systems to complex perceptual channels such as vision and speech. Although some work has already been done in this area, the general problems of interpreting visual images and connected speech are difficult long-term research areas

Current Research Problem Areas

Though we have reached a point where we can develop expert systems that can provide significant assistance within a narrow task domain, the systems which have been developed to date all suffer from several serious weaknesses. In general, there are (1) system development limitations, (2) competence limitations, and (3) use limitations

System development limitations:

- Constructing an expert system requires several man-years of effort from a programmer with a background in artificial intelligence; since very few people have such a background, the number of expert systems currently being developed is small.
- The expert systems which have been developed to date are not at all general; each system is a "single customer", system
- Since the knowledge which an expert system has is collected over time, often from several experts, it is not uncommon for some of the system's knowledge to be inconsistent; there are as yet no good methods for identifying such inconsistencies, let alone repairing them.

Competence limitations:

- Since the knowledge expert systems have is relevant to a narrow domain that is somewhat arbitrarily delimited, the systems sometimes make myopic judgments.
- The systems do not have the ability to check their conclusions for plausibility; thus they sometimes make incredibly naive recommendations
- Since the knowledge the systems have is almost exclusively "surface" (empirical) knowledge, they are unable to infer missing knowledge from general principles; thus their behavior degrades badly when knowledge is missing

Use limitations:

- Almost no expert systems have natural language understanding systems as front ends; as a consequence, users may find expert systems unnatural to use.
- Though nearly all expert systems can provide explanations of how they arrive at their conclusions, these explanations are often not very convincing because they are not tailored to individual users.
- Most expert systems take longer to perform a task than is required by human experts

Research Opportunities

While recent progress in expert systems has led to a number of practical programs, and to a strong interest by industry in this area, expert systems technology is still at a very early stage of development Because we are at such an early stage of development, the single most important research investment in this area is probably to fund basic AI research. This section points out several specific research opportunities likely to lead to increased capabilities and a broadened impact for expert systems in the coming decade.

Knowledge Acquisition and Learning. Given the well-recognized knowledge-acquisition bottleneck, one major opportunity for increasing the power and decreasing development costs of expert systems is in developing new methods for knowledge acquisition and learning

A small amount of research is currently going on in this area, ranging from development of interactive aids for validating, examining and debugging large rule bases, to more basic research on automated learning and discovery of heuristics. The former type of system (e g TEIRESIAS, Davis; SEEK, Politakis) has already been shown to be useful in development of expert systems. Systems of the latter type are still in the basic research stage, and further progress in this direction could have a major impact on expert systems technology

Examples of interactive aids for debugging a rule base include:

> TEIRESIAS system aids user in isolating faulty rule in chain of inferences that leads to incorrect conclusion
>
> SEEK — system collects statistics on rule performance over a database of known correct patient diagnoses, to isolate incorrect rules, and suggest possible revisions

Examples of systems that infer new rules from provided data include:

> Meta-DENDRAL — infers rules that characterize behavior of molecules in mass spectrogram, for use in DENDRAL system for chemical structure elucidation; infers these from given set of molecules and their known spectra.
>
> INDUCE/PLANT — infers rules that characterize plant diseases, given data of symptoms and known correct diagnoses These rules yield expert performance comparable to that attained using rules provided by human experts

Examples of systems that learn heuristics (e g *control* knowledge, as opposed to factual domain knowledge) include:

> LEX -- learns heuristics for selecting among alternative applicable rules, in solving symbolic integration problems, analyzing the solutions, proposing heuristics, then generating new practice problems, etc.

EURISKO — discovers new circuit structures for "high-rise" VLSI circuits, and discovers heuristics for guiding its search for new circuitry structures. This is an interactive system, and has been used in additional domains, such as elementary number theory, and naval fleet design.

While systems have demonstrated the feasibility and utility of computer aids for knowledge acquisition, further progress can have a major impact on development costs for expert systems, as well as on the level of complexity of systems which can be constructed Specific problem areas for near-term research on knowledge acquisition and learning include research of methods for interactively and automatically analyzing and validating an existing knowledge base, and for isolating errors Other promising directions in this area include:

1 Research on methods for inferring new inference rules from problem solving experience For example, in an expert system for interactive problem solving, those problem solving steps provided by the user constitute inference steps that the system might assimilate and generalize into rules of its own for subsequent use
2. Research on compiling "deep" knowledge into more efficient "shallow" inference rules (This is in the context of expert systems that integrate multiple levels of knowledge of the problem domain)
3. Longer term basic research on machine learning, essential to progress on developing shorter-term knowledge acquisition aids Such basic research issues include developing methods for: (a) extending representational vocabulary; (b) learning by autonomously generating, solving and analyzing practice problems; and (c) learning domain knowledge by reading textbooks.

Representation. The representation of knowledge continues to be an area of fundamental significance to AI. By changing representations one can drastically affect the functionality, efficiency and understandability (and therefore modificability) of expert systems

One key subtopic in this regard is the question of *representational adequacy*, i e. the question of whether there is any way to encode certain facts within a language. In the past, research in this direction has led to results such as extensions to predicate calculus necessary to handle default knowledge Representation facilities are another key topic In the past such research has led to the development of convenient specialty "languages" based on frames and semantic nets. Perhaps most important is the actual representation of certain aspects of "naive physics" knowledge, representations of time, space, matter, and causality

Inference Methods. Inference methods are crucial to the expert system methodology so that programs can apply facts in their knowledge bases to new situations. While substantial work has been done on inference, both by logicians and researchers in AI, certain forms of inference of particular significance require further study. Important topics here include reasoning by default, reasoning by analogy, synthetic reasoning (i.e. design), and especially planning and reasoning under uncertainty.

Meta-level Architecture. Meta-level architecture is a new area for research that has considerable potential significance. The goal is the construction of progams that can explicitly reason about and control their own problem solving activity.

The approach here is to view the problem of problem solving control as an application area in its own light, just like geology or medicine. Control recommendations can be expressed in a "declarative" fashion and a program can reason about these recommendations in deciding what to do In this way one can build a system of multiple representation and inference methods, supply it with facts about its goals and methods, and allow it to decide which to use in a given situation. Many traditional knowledge representation techniques, such as defaults and procedural attachments, can easily be expressed as meta-level axioms Key subproblems include the meta-level encoding of standard results from theoretical computer science, and the compilation of programs built with meta-level architecture

Research on User Interfaces. As expert system applications grow into increasingly complex problem solving areas, the importance of high-quality user interfaces will increase as well Progress here involves issues typically associated with man-machine interfaces, such as the need for high-quality graphics, friendly interfaces, fast response time, etc. In addition, the nature of expert systems applications usually requires that a system be able to explain its problem solving behavior, how it reached its conclusions, what inference steps were involved, and whether the problem at hand is one for which is possesses appropriate expertise. For example, an expert system for medical diagnosis and therapy is much more acceptable if it can explain the reasoning behind its diagnosis and therapy recommendations. While current systems have capabilities for enumerating the inference steps involved in reaching conclusions, more sophisticated explanation facilities would greatly improve the acceptability and utility of these systems. One significant opportunity for improving user interfaces (especially for naive users) is to incorporate natural language interfaces into expert systems.

Organizational Problems and Needs

AI has a number of special difficulties because of the ambitious, open-ended nature of its enterprise, and because of the size of the research community compared to the size of the technical problems In addition, AI suffers from many problems that are shared with the rest of science and engineering.

Even restricted NL and expert systems can be very large, and since the success of an approach cannot really be measured until a system embodying it is completed and evaluated, the design-test-redesign cycle tends to be long The development of a full scale system is also a problem

Let us suppose that we want a NL system with a 10,000 word vocabulary; while a portion of a lexicon for the NL system can be extracted from a dictionary, each definition also requires separate attention, since dictionaries do not contain all the information necessary (definitions tend to be circular). If a team approach to construction is used, the team must be small in order to avoid the types of problems encountered in the construction of operating systems – as a programming team gets larger, its members tend to spend more and more of their time talking to each other about standardization, inter-module communication, updates, etc., and produce fewer and fewer lines of code per person per day. On the other hand, if a programming team is very small, it will simply take a long time to produce the necessary code because of individual programming speed limitations. Similar problems arise with the building of knowledge bases for expert systems.

Progress is also slowed because Ph.D. candidates are expected to work independently; the design and building of any significant NL or expert system generally cannot be split very many ways while still allowing suitable academic credit to be assigned to all participants. Once a system has been shown to be feasible, there are further difficulties; universities are not in a good position to develop or support software, since the designer/builders generally leave after receiving a degree, and there are few academic points awarded for cleaning up someone else's system and making it robust.

Industries have special problems as well The largeness of NL and expert systems, and the fledgling state of our current technology makes it difficult or dangerous to promise profitable products on any time scale short enough to be appealing to management. There is also a wide gap between many of the kinds of programs that are produced in academia and the kinds of programs that can become good marketable products. For example, while story understanding programs are considered (rightly) an important current academic topic, such programs have no clear short term market potential.

The comments in this section so far have assumed that we will continue to do research in roughly similar ways to those that have been employed in the recent past, that is, that we will attempt to build AI programs that are "instant adults", systems which are the result of programming, not of learning. There has been a growing interest in learning over the last few years, although only a fraction of the learning research has been directed toward NL and expert systems Certainly there are serious difficulties with the engineering of a system capable of learning NL or expert knowledge from experience (an "instant infant"). For NL, the most serious problem seems to be adequate perceptual apparatus for extracting "important" items from raw sensory data Even if we could devise such a learning program, there is little reason to suppose that it could be programmed to learn language much more rapidly than a human can, which would mean that at least several years would have to pass before we could judge whether or not the system was adequate And, of course, the chance of getting everything right the first time is close to nil Nonetheless, this seems to be an important avenue to explore, to hedge our bets, to further cognitive science and perhaps to find a compromise position ("instant five-year-old"?) that would represent the optimal long-term route to fully general NL systems. Further reasons for emphasizing learning include the observations that the only language users that we know are built through learning, and that we continue to use learning as adult language users, so that we probably need learning of some sort in our programs anyhow

For expert systems, a key difficulty is that human experts are generally not very good at explaining the basis of the expertise. We are very far from knowing how to design programs that could build a suitable knowledge base and body of inference rules by simply watching problem situations and the expert's solutions for them. Again, such a system would have to have sophisticated ways of judging what items from its experience were important, and of inferring the nature of the knowledge of the expert from these items. This seems impossible without beginning with a highly structured system, about which we currently have very few concrete ideas

Another practical problem is finding good application areas, that is, ones where NL or expert systems can be truly helpful, where the domain is well-circumscribed and well-understood, where the tools that we now have are capable of conquering the problem, and where typing is possible and acceptable as an input medium

Finally, there is an acute shortage of qualified researchers. Since most actual application programs are likely to be the result of development by industry, it seems desirable to encourage more industrial effort However, taking researchers away from universities reduce the number of faculty capable of supervising graduate students and carrying on much needed basic research

Manpower

AI shares a severe manpower shortage with the rest of computer science and computer engineering. However, in AI the problem is compounded because a "critical mass" of researchers seems to be essential to carry out first-rate research. To be concrete, let us list some of the requirements for doing NL or expert systems research

Software support - - Since AI uses languages and systems software packages and often machines that are different from the rest of computer science, separate systems people are extremely important At least one, and preferable two or more are needed Implementation is essential: ideas cannot be tested without implementation, and many ideas result from the experience of implementation The day of the single-researcher tour-de-force system has largely passed. Because implementation is a lengthy and costly process, it is important to argue before implementing. In arguing, the generation of new paradigms is important a re-implementation of previous ideas is of little value.

Within NL, a modern system should have a user interface, a parser, a semantic interpreter, a knowledge representing and retrieving component, a discourse plan evaluator, a speaker modelling component, and a language generator. In addition, if it works on a data base or knowledge base of some kind, there is a further need for support. There are 6-7 different areas here, and even if each person is expert in three, there is a need for four or five people. The story is similar for expert systems: an expert system must have a large knowledge base, which generally requires at least one expert and one knowledge engineer. An expert system also requires a user interface, which involves both input understanding and output generating components. Again the minimum critical mass is four or five researchers. Using senior graduate students is possible, but turnover after graduation causes serious problems in keeping a system working. Using these criteria, only three or four US universities have a critical mass in NL, and a similar number have critical mass in expert systems.

People who want to do AI research must also be cognizant of fields outside computer science/computer engineering: NL people ought to know linguistics and possibly psychology and philosophy of language; expert systems people must know about the knowledge area (medicine, geology, molecular biology) in which their systems is to be expert, vision people need to know about neuro-physiology and the psychology of perception, and so on. This means that effective critical masses may be even larger, and may only be possible in settings which can provide the right support outside of computer science/computer engineering

Leaving the critical mass issue, there is a critical manpower shortage at universities, especially for research and for graduate research supervision. Few new Ph D.s are being produced, and computer science suffers from a singularly high emigration to industry Part of the problem is teaching loads. In a time of rapidly expanding demand for education in computer science/computer engineering, both for specialists and for those who want computer science electives, teaching loads are likely to be high. This is compounded by heavy graduate advising demands Of the six faculty members in our group, one supervises twelve graduate students, two supervise ten each, one supervises six, and one supervises two However, the one who supervises two also supervises eighteen full-time professional staff members and the one who supervises six has a relatively heavy teaching load. When the need to write proposals for the substantial equipment and research funding needed to support serious research are added, it is clear that faculty in AI may well have difficulty fitting in time for research. Graduate students have in some cases decided not to go into academia specifically because of what they have observed of their advisor's experience

Industry too has problems, many not specific to AI. There is such a shortage of qualified personnel that nearly as many non-Computer Science (CS) majors as CS majors are hired for CS positions. CS education at universities also varies wildly in quality, especially at minor institutions, and there is no easy way (short of some form of profession certification) to judge the competence of prospective employees. No doubt, many of the problems in generating high quality software stem from poor education.

Equipment

AI research requires extensive computer facilities, with different types of computers than those used for large numerical tasks. There is a wide gap between the top few universities and research labs, and the rest of the universities doing AI research. The top several AI centers now have one powerful personal computer (usually LISP machines) for every three or four researchers (graduate students and faculty) as well as a number of larger time-shared machines, ARPANET connections, and other special-purpose equipment

If reasonable AI research is to be done elsewhere, much more equipment must be made available. More equipment is also critically important to keep new Ph.D graduates in academia: a number of industrial labs are at least as well equipped as the best universities. Graduate students are unlikely to want to go to a place that has computer facilities that are inferior to those they are accustomed to.

AI researchers at universities which do not have software support groups are now constrained to use relatively narrow range of equipment types in order to be able to use shared software from other locations. The manufacturers of this preferred equipment (DEC, Xerox, Symbolics, LMI Inc., and a handful of others) have not donated much equipment in recent years; the few exceptions have been universities that are already relatively well-equipped. The future looks a little brighter, since a fair body of AI software can now run under UNIX, and many manufacturers either offer, or plan to soon offer, machines (especially based on the Motorola 68000 chip) that can run UNIX. Maintenance funds are also critical; there have been cases where industrial gifts have been turned down by universities because maintenance was not included, and the universities had inadequate budgets to pay the maintenance costs.

An Assessment of Resources Required for AI Research

AI's main needs are manpower and equipment. Except for very limited special purpose applications, AI systems require relatively large program address space and relatively large numbers of machine cycles When done on the time shared machines in general use (e g. Digital Equipment Corp VAXes and DEC-20s), research is clearly hampered by a shortage of machine power. LISP machines are much better, but some (e.g. Xerox 1108's at $32K each) have rather small address space, and the rest tend to be quite expensive (more than $75K) if dedicated to a single user. Very roughly about $25K-$35K in additional funds per researcher is needed to provide at least a semblance of the best environment for AI research; anything less will mean that progress is slower than it could be. In this estimate, we have assumed that a number of researchers will get more expensive machines (Symbolics 3600, LMI Nu machine, or Xerox Dorado), and that shared facilities, terminals, modems, printers, mass storage, etc must be purchased also. Including graduate students and research programmers, as well as faculty researchers, there are now probably 150 people doing work on NL and expert systems in universities. This brings the total expenditure needed to properly equip these researchers for maximum progress to about $5M. About $50K-$60K is required to properly equip each new faculty member.

Manpower needs cannot be solved quickly by simple infusion of dollars. Money that is spent on university equipment will probably do the most good, because it can help speed the research for graduate students, and it can also make universities relatively more attractive places to induce faculty members to remain and new graduates to choose in place of industry. This will in turn accelerate the production of new researchers. At the best, we are still likely to have a serious shortage of AI researchers for the foreseeable future: given that there are fewer than 25 faculty members who each graduate about 0 5 Ph D. students per year, we can expect for the near future only 12 or so new NL Ph.D s per year. In expert systems, the situation is a little better, with about 30 faculty nationwide, but there will probably still be no more than 15 new Ph.D s per year over the next ten years. Probably only about half of the new Ph.D.s will go to universities, and it will be five years before the new faculty produce their first Ph D. students. Thus the manpower situation for the next ten years is likely, in the absence of any massive intervention, to leave us with fewer than 100 NL faculty and about 100 expert system faculty members nationwide at the end of the ten year period

As mentioned above, a possible way to increase our AI capability in the shorter term would be to encourage a cross-over of faculty from areas such as linguistics or development psychology to AI NL research, and from various fields of expertise (e g. medicine, geology, etc.) into expert systems research. There are already incentives for researchers with suitable backgrounds, since funding in many other areas (e.g. linguistics) has generally been cut along with social sciences An infusion of researchers from these areas may have possible long-term benefits so AI NL research above and beyond providing more manpower; it is our belief that in order to truly succeed in producing general, robust NL systems, we must develop a far deeper science of human language unerstanding and language development, and it seems clear that expert systems research requires humans processing the specific expertise to be part of the effort.

It might also be helpful to provide graduate fellowships, though the most serious shortage seems to be supervisors of research, not interested students nor money to support the students.

Recommendations for the Next 5-10 Years in Natural Language

0. Don't expect too much; the number of researchers in the field is really quite small, and the size of the task of understanding is enormous.

1 Continue a broad range of basic research support; there is still a shortage of science on which to base NL system engineering

2. Increase funding for equipment Adequate equipment is essential if researchers are to produce working systems, rather than just theoretical advances Such funding has many benefits: Increasing available compute power can dramatically cut the amounts of time and effort required to produce running systems, since easy-to-write though computationally expensive systems can then be considered Squeezing a large program onto a small machine can be very time-consuming. If programs take too long to run, programmers start to work on speeding them up instead of working on increasing their range of competence Making modern, powerful equipment available to universities will help them retain faculty. Despite the apparently high initial cost compared to salaries, money spent on hardware is likely to be a good investment.

3. Encourage resource-sharing, by funding specific research groups to develop supply, and maintain a common body of research tools, for example AI programming languages, natural language parsers, knowledge representation systems, "empty" expert system (i e. reasoning and knowledge base access portions of expert systems with domain-specific knowledge removed), and programs for transforming programs from one language or operating system to another. To some degree, this recommendation is already being followed, and has accelerated research progress.

4 Create and encourage development groups in industry and military labs, and encourage increased contact between such groups and university and industrial basic research laboratories. Universities are particularly ill-suited for developmental efforts, since there is a high turnover of key system builders, making it difficult to support application systems In addition, we need all the effort on basic problems that can be mustered Development could be handled by groups that have more traditional software backgrounds; once feasibility has been demonstrated, AI systems often look a lot like other programs. Possible incentives could include tax breaks and jointly funded university/industry research and development efforts, though the latter would have to be designed judiciously to avoid

waste and mismatched expectations and capabilities

5. Encourage industrial research laboratories to help be advising Ph D. research whenever possible. Such cooperation can benefit the laboratories by providing relatively low-cost, high-quality staffing, and can help increase the size of the U.S. research community at a faster rate. This recommendation can work; SRI International and Bolt Beranek and Newman, Inc , among others, have successfully functioned as Ph D. research supervising institutions for a number of years.
6. Maximize faculty research and research supervision time by providing partial academic year salaries as well as summer salaries
7. Institute three- to four-year research initiation funding, including equipment funds, for promising new graduates who agree to stay at universities Hewlett-Packard has already undertaken such a program
8. Encourage the design of novel supercomputer architectures that take AI needs into account; current supercomputers are number crunchers that are of little use or interest to AI, though AI badly needs computers with greater power. AI researchers have begun to design such machines at a few locations. AI people have had some success at computer design (e g LISP machines), but it would be desirable to have groups that specialize in computer design involved with such designs Japan, in its 5th Generation Computer effort has already undertaken such a goal, and it seems quite possible that even a partial success in their effort can cause serious erosion or loss of the U.S. high tech edge.
9. Encourage cooperation between AI NL research, and the traditional fields interested in language (linguistics and psychology) This can serve the purpose of aiding the building of a science of language and cognition, and can also provide a more rapid increase in manpower resources than is possible through the training of new researchers only This recommendation may happen anyhow, since "social science" funding has been cut dramatically, outting pressure on many of the kinds of people that could help the AI NL effort To make this work, a fair amount of re-education, especially in AI, computation and programming, will be needed along with some re-education of AI people in linguistics and psychology, to build a common knowledge base