CHALLENGES

Susan Esptein
Rutgers The State University
Piscataway, New Jersey

12 October 1982

ABSTRACT

It is the intent of this article to push beyond the horizon of "current survey and future work," with a set of mildly outrageous long-term challenges. The targeted areas are expert systems, problem solving and learning.

Somewhere between AAAI-82 and science fiction, our future research projects lie. If we are not to be mired in the pedestrian, we should reach out beyond the traditional. Not long ago, a proposal to develop heuristics (Lenat, 1982) or follow strategic instructions (Mostow) would have seemed mildly outrageous. This article is intended as a set of equally unreasonable challenges, a focus for the future. Expert systems, problem solving and learning are particularly amenable to such a focus. The categories are neither mutually exclusive nor collectively exhaustive.

Expert Systems

Hypothesize an expert system S which represents knowledge about its domain D in a semantic net. (The examples in boldface below are output from an expert system for the domain of elementary calculus). The nodes of the net represent relationships among those objects. S deliberately models the neural net of human associative memory. Thinking could be categorized as an exploration of this net. A state-of-the-art expert system (e.g., UNITS, MRS, KL-ONE) is merely an *idiot savant*, a computational wizard of impressive speed and accuracy, who lacks the ability to:

- ☐ probe the knowledge base
  S has adequate data, but lacks *procedural* ability. We ought to be able to retrieve data about D without reference to a specific case. Thus we might ask

  Are differentiable functions always continuous?

  and expect some research to take place. S could serve as an encyclopedia or a tutor.

  Consider the following example...

  or

  The tangent function has infinitely many points of discontinuity, yet is everywhere differentiable.

  CAI techniques might well improve the presentation, so that the output from S would be clearer, and better organized than the input had been.

- ☐ Aid in theory construction
  Once again, without reference to a specific case, S ought to be able to test the likelihood of an input theory. Responses might be probabilistic statements, reasoned rejections, or even counterexamples. For the hypothesis "continuous functions are everywhere differentiable" the response might be

  The function $|x|$ is a counterexample

- ☐ Exploit implicative structure
  Each subset of the nodes has a generated subgraph which may be characterized as the implications deriving from it.

  What are the properties of $f(x) = \text{secant}(x) + \log(x)$?

  Each subset of the nodes can also be traced backward to a set (or sets) of supporting or causal nodes.

  What do the tangent, arctangent and secant functions have in common, and to what is that commonality attributable?

  Each subset of identically labelled arcs characterizes a relation in D.

  How does integrability relate to differentiability?

  Finally, it may be possible to prune the net by determining minimal set-of-support data. Net pruning must be domain-dependent, reflecting both safety and cost-effectiveness.

- ☐ Explore
  Bayesian or logical contradictions may be inherent in the data used to construct S.

  It is inconsistent with my understanding of Integrability for that to be correct. Please explain.

  S should be able to explore and criticize its own knowledge base.

  I would appreciate an example of a discontinuous, integrable function.

  Reference to an experiential library of example and the ability to construct counterexamples would be crucial here.

  I believe I have constructed a counterexample to theorem 3. Please consider the following...

- ☐ Monitor reasoning behavior
  Long-term memory could lead to suggestions for modifying the net

  Examination of the points of discontinuity has proven effective in the past so...

  Such changes would be based on experience, a crucial component of expertise. (Lenat, 1979)

- ☐ Formulate a theory
  If D has an underlying theory, the net can be presumed to reflect portions of it. Thus it is not unreasonable to have S postulate hypotheses for its observations and request experiments whose results will confirm or deny them. For a mathematical domain, such an experiment could be to call upon a theorem prover; for a physical science, the experiment would occur in a laboratory.

  I have postulated the following theorem and am attempting a proof...

Problem Solving

Many of the classic problems (missionary and cannibals (Amarel, 1969), Tower of Hanoi (Amarel, 1981), etc.) have been

exhaustive explored. A cohesive body of knowledge about problem solving will still require:

- A set of target problems
  There is now a committee actively soliciting unsolved mathematical problems which would be amenable to solution by theorem proving (CADE). An extensive set of challenges to AI problem solvers and a taxonomy for those problems would be an important organizational focus. Any available solutions or analyses, such as the recent work on Rubik's cube (Korf), should be included. This collection must ultimately be an important component of an expert problem solver.

- Mathematical domains
  Despite Slagle's (Slagle) early success with integration and Lenat's (Lenat, 1977) more recent success with set theory, little ongoing work exists in mathematical domains. Yet mathematics forms a natural bridge between the deceptive simplicity of toy problems and the mysterious complexity of the physical sciences. Certain areas of mathematics (e.g., number theory, graph theory, topology) offer the definitions and relationships of toy problems as well as the uncertainty about underlying principles inherent in the physical sciences. Support should come from Piaget's work and yet-undone protocols of mathematicians at work.

- Representational transformations
  Amarel (Amarel 1969, 1979) and Anzai and Simon (Anzai) have amply demonstrated the power of representational changes in problem solving. AI lacks an understanding of the origin and nature of these transformations. Particularly significant are the choice of data structures, the sequence of the representations and the experiential bases for selecting a representation.

- Analogy
  Carbonell's (Carbonell) work in analogy in an important beginning. There is increasing evidence from the history of science (Darden) and strong intuition in AI that analogy is crucial in theory formation. Visual analogs should be particularly fertile. For these AI will require paradigms from work in both vision and psychology. This author believes that many representational transformations may be explainable by mathematical and/or visual analogy between data structures. In addition, the diaogue between an expert and an expert system-in-development could be facilitated if the program were to draw analogies between the current domain and others it had previously encountered.

Learning

Some of the earliest AI work hypothesized the modeling of a large infantile brain. This brian would then learn from a carefully structured environment, rapidly and exhaustively acquiring factual and procedural data. More recently, expert systems develop a data base by interaction with an expert, a primitive form of learning. The following characteristics of human learning are appropriate challenges for AI:

- Retention in associative memory
  Humans gradually link each new object or belief into an elaborate semantic network. Both psychology and neurophysiology are important here. These links apparently include aural and visual, as well as contextual cues. Forgetting and detecting anomalies should be important here.

- Self-awareness
  Humans monitor their own progress and are able to differentiate on a spectrum from rote learning to the satisfying grasp of a concept. Key here are thorough establishment in the semantic net and continual attempts to analogize. This meta-learning evolves in humans as both domain-specific (e.g., academic discipline) and domain-independent (scientific method) techniques.

- Inquisitiveness
  An artificial intelligence should rest only for occasional maintenance. In the absence of outside stimuli, a learning machine should be able to formulate questions for its tutors based on what it observes as gaps in its semantic net in the absence of tutors, a learning machine should be able to hypothesize and test more abstract formulations. LEX (Mitchell) is a admirable step in this direction.

- Application
  A machine which has learned a concept must be able to utilize it in a variety of ways. Thus procedural knowledge becomes an important kind of learning. Of particular interest should be the link between the identification of an object/fact and its origin, the ability to formulate examples and counterexamples, and the ability to synthesize efficient combinations.

Conclusions

This author has noted among AI researchers an increasing facility in generating plots for science fiction stories. Such fantasies are not the stuff of dreams, but reasonable extrapolations into the future. I wish to thank AAAI-82 and Saul Amarel, Rusty Bobrow, Lindley Darden, Doug Lenat, Marvin Minsky, and Tom Mitchell as catalysts for the challenges outlined here.

References

Amarel, S. (1968) On Representations of Problems of Reasoning about Actions. Machine Intelligence 3:131-171.

Amarel, S. (1981) Problems of Representation in Heuristic Problem Solving; Related Issues in the Development o f Expert Systems. Tech. Rep. CBM-TR-118, Computer Science Dept., Rutgers University.

Anzai, Y., H. Simon (1979) The Theory of Learning by Doing. Psychological Review 86:124-140.

CADE (Conference on Automated Deduction) resolution, June, 1982.

Carbonell, J. (1982) Experiential Learning in Analogical Problem Solving. AAAI-82:168-171.

Darden. L. (1982) Reasoning by Analogy in Theory Construction in PSA 1982, volume 2, Philosophy of Science Association, forthcoming.

Korf, R. (1982) A Program that Learns to Solve Rubik's Cube. AAAI-82:164-167.

Lenat, D. (1976) AM: An Artificial Intelligence Approach to Discovery in Mathematics as Heuristic Search, Ph.D. dissertation, Stanford, 1976.

Lenat, D., F. Hayes-Roth, P. Klahr (1979) Cognitive Economy. A Rand Note. N-1185-NSF.

Lenat, D., W. Sutherland, J. Gibbons (1982) Heuristic Search for new Microcircuit Structures: An Applications of Artificial Intelligence. AI Magazine 3:17-33.

Mitchell T., P. Utgoff, R. Banerji (1982) Learning by Experimentation: Acquiring and Refining Problem Solving Heuristics in Machine Learning, J. Carbonell, R. Michalski, T. Mitchell.

Mostow, J. (1981) Mechanical Transformation of Task Heuristic into Operational Procedures, Ph.D. dissertation, CMU, 1981.

Slagle, J. (1963) A Heuristic Program that Solves Symbolic Integration Problems in Freshman Calculus, JACM 10:507-520.

A NOVEL INTERACTION BETWEEN AI AND PHILOSOPHY

Giuseppe Trautteur
Istituto di Matematica
Universita di Palermo

1. Most philosophical ideas and systems can be regarded as purportedly veridical descriptions of the *status quo* of some particular aspect of reality or of reality *in toto.*

Examples particularly relevant to the following are: the scholastic ontology and theory of knowledge, British Empiricists cognitive constructs, Leibniz' Monads, Kant's transcendental aesthetics and analytics.

Most of these descriptions take the form of processes where something "becomes" something else while maintaining more or less large amounts of self-identity. Or something implies something else or exerts some action either on something else or on itself in reflexive ways. Notions such as the *concept*, the *idea*, the *sense-data*, the *category*, the *thought*, the *subject*, the *substance*, the *act*, etc. are never formally given as basic unexplained terms on which to build the philosophical edifice, but, on the contrary, are often introduced, discussed, explained and allegedly shown to constitute or support a coherent and meaningful edifice on the strength of an informal but semantically loaded, argument. Thus, lacking a full fledged epistemological concern, which does not seem to have penetrated any philosophical system, all systems are naive in the sense of being based more or less straightforwardly on common sense understanding of natural language.

In what follows, I intend to restrict the focus to those philosophical chapters connoted nowadays by "cognitivism" which were known in the old days as Logica Maior, parts of Metaphysics, Psychology and Gnoseology and which can be taken as a theory of "Common Sense."

A philosophical system (by which I mean one of the historically identified sets of ideas just mentioned) can be exploited in at least three different ways:

1. as a bona fide description of what is the case;
2. as a set of statements, ostensibly conveying ideas, to be studied for their historical (or otherwise, but not exclusively *per se* as sub i.) interest;
3. as an yet untested hypothesis about the actual working of human understanding on which to perform experiments of simulation, corroboration, falsification, validation, etc.

[1]Partial support of the Italian CNR under contract no. 80/2199.07 is gratefully acknowledged.

2. Artificial Intelligence (AI) may be seen as the continuation of Cybernetics in the western sense of the word. As such it seeks an explanation and an understanding of all phenomena pertaining to complex organizations and particularly to the symbolic activity of the human being. The real cornerstone of this endeavor is the physical existence, not the mere possibility, of a large, fast and complex algorithmic device. This is so because mechanical explanatory phantasms such as, those of Descartes and de la Mettrie, can now be considered not just more or less correct metaphors, but hypotheses testable *in copore vili.* The kind of reduction entailed by this situation is quite at variance with the classical reductionistic scheme. There the *explanans* was convincingly and patently connected to the *explanandum* at least in the simplest situations such as, say temperature and the microscopic distribution of velocities. Here we are not (yet?) in a position to directly correlate the so-called higher mental functions with their physiological *explanantia.* Instead we are developing systems vicarious to the functions of the brain in what can be seen as an unformalized quest; one where the *explanans* is not quite extant, but resides in the rigorous but informal understanding of the searcher: a situation similar perhaps to that of eighteenth century physics when there was nothing to which the mechanical concepts could be reduced. On the contrary notions such as momentum, energy and their phenomenology were created *ex novo* to become themselves, in centuries to come, the "intuitive" basic *explanantia* of other disciplines. AI therefore treads a methodological compromise where:

1. a direct experimental assessment of mental phenomena is sought after in alleged or tentative duplications of mental phenomena;
2. the tacit presupposition that mental phenomena are algorithmic, and thus material is maintained under the form of (some version) of the Church Thesis;
3. the possible success of a Turing test could be the premise for a further understanding of the mental (heuristic aspect of AI).

One should also not forget that the kind of cognitive object AI is after is nothing less than the human mind itself. Here - it has often been suggested - a knowledge of a totally objective nature is perhaps impossible, albeit no formal limitative theorems have been discovered yet, despite arguments of the sort offered by Lucas (1). On the contrary, it might be that the knowledge of the human mind will resort to or include in some constitutive way the acknowledgement of an intentional "thou," or at least of an intentional stance, to follow Dennett (2), whenever the consciousness becomes an issue. And the whole epistemic outlook would change, the reduction becoming a witnessing. Nonetheless, even if the *res cogitans* is expelled, the consciousness obviously will not be. One message of AI research, coupled with a careful meditation of the extant limitative theorems, seems to be that Church Thesis mechanistic structures might not be strictures after all and that genuine understanding (consciousness) is possible within their bounds.

3. Let us then assume that research in AI strives to duplicate human mental behavior on the basic assumption that it be algorithmic. Moreover, this activity is often perceived as a novel explanatory paradigm where the analogical power of a program which "works satisfactorily" is (at least a first step towards) a satisfactory intersubjective understanding of the human mind. But:

1. what is it to be "satisfactority working" for a program;