AI Magazine Volume 6 Number 2 (1985) (© AAAI)

AI, Genetics, and Applied Epistemology

Editor:

In principle, AI science can be divided into two distinct areas of inquiry:

(1) genetic engineering, which seeks to affect the evolution or effect the synthesis of intelligent behaviors that organic artifacts display; or,

(2) computer architecture, which seeks to discover or invent epistemic resources and their means of refinement, yielding the knowledge processing facilities and functional capabilities of AI machine artifacts.

Moreover, the scientific framework of AI can consist of two distinct kinds of theories and methods

(a) extensional, descriptive accounts of what actually is known, how knowledge actually is acquired, and what has been done with it, as matters of historical fact, and,

(b) intensional, nomological accounts of what there is to know, how it could be known, and how knowledge could be used, each in principle, apart from actual human practice, expert or not.

Currently, the program for AI science is (at least) predominantly descriptive, extensional, and historical. *Scientific Knowledge* (Fetzer 1981, D. Reidel) illustrates that an idealized conception involving intensional language and logic is required, however, if the nomological character of scientific laws is to be represented in a formal system. The probabilistic causal calculus C that Fetzer advances can accommodate precisely the degree of epistemic adequacy (*i.e.*, "robust" knowledge processing) that is ultimately sought in the construction of AI machine artifacts, thus offering a formal epistemic architecture of language and in logic which other formal systems cannot afford. Recent papers by Jim Fetzer and Terry Rankin (1984) and Terry Rankin (1984)* strongly suggest that a normative program of *applied* epistemology can indeed be significant in AI science (for AI computer architecture, at least), offering many important advantages and major advances that cannot be achieved via current programs of AI research and development.

Properly conceived, formal epistemic architectures can facilitate both deduction and induction, for example, as complementary modes of monotonic and nonmonotonic inference, respectively. Moreover, other problems of material conditionality and transitivity can be resolved through intensional semantics, and the implementation of normative theories of knowledge for AI machine artifacts reflects precisely the sort of systematic and comprehensive framework that AI science currently does not have. Despite Bierre's claims that formal systems have outlived their usefulness, quite the opposite is true: formal systems are indispensable for ensuring epistemic adequacy in the construction of AI machines. The problem at hand is not to abandon formalization in favor of ineffable heuristics, but rather to idealize the process of formalization away from its extensional and descriptive inertia and toward an intensional and nomological conception of both ordinary and scientific knowledge. Failure to do so will only confine AI indefinitely to its current status of *ad hoc* analyses and methods, lacking the systematic procedures and comprehensive theories which distinguish a genuine science.

Terry L. Rankin
University of Georgia
Advanced Computational Methods Center
Athens, Georgia 30602

Response to "The Professor's Challenge"

Editor:

Pierre Bierre raises some interesting issues in his recent article, "The Professor's Challenge." In adopting the "egg" point of view, he is really arguing for a bottom-up approach to knowledge acquisition. His "principle of prerequisite knowledge" captures the essence of bottom-up reasoning: if a higher level symbol is going to rely on a lower level symbol in some way, then the lower level symbol should already exist at the point that the higher level symbol comes into existence.

The benefits of a top-down approach to software development have been discussed extensively, and some of the arguments also apply to the notion of top-down knowledge acquisition. In particular, a strong case could be made that a thinking machine must be interested in something before it knows what knowledge to acquire. Such interest amounts to a top-down hypothesis of sorts—that is, the thinking machine acknowledges that there is something to be known at the same time that it admits that it does not yet know it. By breaking down its goal (*i.e.*, the thing it is interested in) into pieces, it can proceed to acquire them one by one

Bierre's "religious conviction" on the side of the "egg" camp is an extreme position. In taking this position, he jumps into the sea of relativism where there are no external laws—a symbol is meaningful only if an individual thinking machine learned it for itself. His position does not allow any of the benefits of *a priori* knowledge or a top-down approach to learning.

However, Bierre's argument points out how relevant philosophy is to the design of knowledge acquisition systems. (Perhaps all of the work that philosophers have been doing over the centuries, which has often been characterized as profoundly unpractical, will suddenly assume great practical importance in the design of knowledge acquisition systems.) The "chicken" position discussed by Bierre corresponds closely with the "realist" idea of perception, and the "egg" position corresponds to the "idealist" idea. I think a strong case can be made that neither position is correct by itself, but that any thinking machine must alternate between the two in order to acquire knowledge.

Where Bierre seems overly ready to give up on "logic and deductive inference," other researchers are expanding the frontiers of these disciplines. Specifically, there is another approach to knowledge acquisition stemming from the philosophical notion of dialectics, in which alternative "theories" or explanations of external phenomena are maintained by the system until one of them can be demonstrated to be clearly superior. (See Kuhn (1962) and Popper (1967) for discussion of dialectics in the development of scientific theories.) This approach has been developed theoretically by Kowalski and Bowen and is now being pursued in practical systems by Miyachi at ICOT.

Bierre speaks of an intelligent system which interfaces to the world exclusively through "sensorimotor data streams," presumably in place of current systems which rely on machine-user interactions, or, worse yet, on some programmer typing rules directly into the innards of the machine. But this distinction may just be one of level. Isn't it reasonable to consider a computer terminal or a joystick as the "sensor" (*i.e.*, a way of getting information from the outside world) of an intelligent system, in which case all interactions with the user while the program is running become its "sensory data stream"? Is there something inherently purer about the output of an A-to-D converter than the output of an input parsing routine? My own position is that there is very little difference between the two kinds of output when they become input to an intelligent system. To attempt to take refuge in the low-level silicon (as the source of intelligence) is futile, because, ultimately, the only thing an intelligent system can do is to reflect the subjectivity of its architect.

Bierre's final controversial point is that AI needs a limit to its domain; in other words, a concept of the "totality of knowledge." This point is probably not brought up often enough because it is so humbling to those involved in AI research. At the same time, it is worthwhile to remember that other disciplines, philosophy included, would also benefit from such a concept when it becomes available.

Perhaps this is the area where AI has an important practical contribution to make to philosophy, or where AI and philosophy will enrich one another through their interaction.

John Malpas
Pulsetrain
747 Greenwich Street
New York, NY 10014

Bierre Responds

Editor:

John Malpas raises a very interesting question: What difference does it make whether we interface intelligent machines to the outside world through continuous-running sensory data streams versus formal language objects and commands?

The difference lies in the machine's usefulness for the transmission of knowledge. Let me relate a true story that happened recently as an example. I was having an expert system for diagnosing respiratory disorders demonstrated to me, and the first question asked of the user was: `What is the quality of the stridor?`

Stridor? What's that? The expert system was in no position to explain what it meant by stridor. The creator of the knowledge base knew, and was counting on the person at the other end to already know, too. It turns out that stridor is a sound made by congested lungs... you have to hear some before you really know what the word refers to. In the expert system, stridor is manipulated as a literal symbol; all that is known about it is the ASCII bytes that somebody typed in when they first defined it as a variable.

There is nothing unique about expert systems in this regard. *All* computer systems are "face value" symbol processors, as Ed Feigenbaum likes to point out. Mr. Malpas shows excellent intuition in his suggestion that, from the computer's point of view, input is input is input Indeed, the computer has no foolproof way of knowing where its input data is coming from All it knows with certainty are the bit values pouring in.

With this fundamental limitation of computers in mind, *i.e.*, their computations are ultimately grounded in primitive tokens that enter the system undefined, the question then becomes how can we build knowledge systems where the primitives are *equally meaningless* outside the computer as they are inside it. In other words, if computers are inherently stupid about what they are computing about, let's figure out the level at which humans are inherently stupid about what we are computing about, and then devise a human-machine interface that operates on this level of communication. What level would that be? I draw your attention to the primitive "states" appearing at the remotest periphery of the cochlea, retina, and other sensors.

When the programmer of the respiratory disorders knowledge base decided to put stridor into the system as a pertinent variable, stored upstairs in her head and those of other human experts familiar with lung ailments was knowledge about what stridor sounds like, knowledge that cannot be transmitted by the expert system because its primitives are coined at too "high" a level. But when a medical student learns about stridor firsthand in the hospital ward, the primitive states being "entered" at his or her cochlear and retinal surfaces don't have any meaning known to somebody outside. They carry no meaning whatsoever. The fact that humans learn and communicate in terms of sensory states that have no semantics accounts for our ability to transmit knowledge amongst each other without running into the problem the expert system story illustrates

From this perspective, it becomes clear why sensory interfaces are not just another way of getting information in, but an indispensible element in the grand scheme of intelligence. They make it possible to communicate in a "language" whose tokens require no further explanation... their informative value is everywhere in the network limited to "face value." It is precisely the ability to trace knowledge back to its face value, sensory roots that 1990s knowledge systems will need in order to become useful as learning and teaching assistants.

This goal will require that human-computer interface designers and library-of-the-future engineers adhere to the broad principles of knowledge transmission being currently developed to explain how humans communicate with each other. Interested parties may obtain more information on transmission of knowledge theory by contacting us.

Pierre Bierre
Clairvoyant Systems
1921 Rock Street
Mountain View, CA 94043

Knowledge and Power

Editor:

This letter was inspired by a disquieting conversation with a retired military physicist who wondered whether a valid analogy could be drawn between the current state of expert systems and the very early days of research on atomic weapons. Starting from the premise that knowledge is in fact power, he pointed out that many kinds of knowledge are restricted to a small number of people, survivors of arduous apprenticeship. He went on to wonder whether making such knowledge available to many people without the concomitant period of reflection and appreciation might not entail unforseen, and grave, consequences. It seems to me appropriate that as creators and developers of intelligent computing, the AI community consider two interrelated questions:

To what extent and in what ways is knowledge power?

The promise of expert and natural language systems is in making knowledge more accessible, and more available for practical application. This can be seen as such a natural continuation of the process which began with the invention of writing that it is beyond being labeled good or bad—it is just the way it is. But that does not absolve us of responsibility for the question of what kinds of knowledge, in whose hands, yield what kinds of power.

Let me express my fear, in the hope that members of the AI community can respond with good cause to allay my fear. Intense military research and development, combined with the drive to host expert systems on microprocessors, will make sophisticated military powers available to many groups which today have no such capability This capability will be used to magnify the distance between groups in power and those out of power. Increasing distance will cause the out of power to perceive fewer and fewer options other than terrorist violence. Increasing separation of the haves and the have-nots means increasing fragmentation, leading to increasing polarization. We see the exacerbation, not the healing, of the wounds which exist within and between societies.

The history of technology is in general a history of increasing distance between the wealthy and the poor. The export of technology (with the possible exception of medical and agricultural technologies) to the less developed countries (LDCs) has in almost all cases increased the distance between the controlling groups and those who would replace them. Expert military systems and smart weapons, based on inexpensive portable or embedded micros, seem to me to have the ironic potential to both further cement existing power structures, and in the hands of enemies of current regimes, make more effective by far terrorist attacks on these same structures: a vicious cycle which feeds the urge toward totalitarianism. It would be a sad irony if AI applications, instead of fulfilling their marvelous potential for the democratization of technology, lead us further away from democratization of anything

If the above fear is reasonable, what is our responsibility as a community of researchers regarding the development and employment of military AI technology?

On the one hand, we might reverse in this special case the well established pattern of funding driving the output of research—we could, in general, refuse to do so much military R&D. But not only is this extremely unlikely, it has two important flaws. First, the point of much of our U.S.

technology is the avoidance of war Whatever one considers a "legitimate security interest" to be, precise and reliable information-gathering, analysis, interpretation, and communication are critical to maintaining peace. Second, it ignores the possibility raised by Freeman Dyson that the extremely focused and targetable power of intelligent weaponry will render large-scale nuclear, chemical, and biological weapons obsolete.

But on the other hand, to dive in unreservedly to military AI applications may tend to bring about the scenario stated above.

Clearly, I am offering questions rather than answers. I am eager to hear the thoughts of others in the AI community. With many of us poised on the edge of Strategic Defense Initiative ("Star Wars") work, the time is ripe for discussion

Jim Kornell
921 North Alisos Street
Santa Barbara, California 93103

Editor:

I wish to inform you and your readership that in the Spring issue (Vol 6 No. 1) there has been a number of pretty unpardonable mistakes, omissions and unjustified editing in the version of my address, "Knowledge and experience in artificial intelligence". This is quite apart from the very many gratuitous splittings of integral sentences into parts, changes of spelling, etc.

The following are some of the worst examples:

The last sentence of the second paragraph of Page 41 has been converted into gibberish by making the subject of the mail clause "modelling a simulation..." instead of "a simulation of a real world".

In the last sentence of the seventh paragraph of the same page, not only has the meaning been seriously distorted, but a good joke has been converted into a remark that can only be described as that of a moron. It should have read, "As I believe Raphael put it, they found that the best model of the world <u>was</u> the world—<u>and</u> cheaper!

In the fifth paragraph of Page 42, the editing has resulted in hiding the fact that my text referred to a specific book of Freud, namely "The Interpretation of Dreams".

The following examples are perhaps a little less serious, but some of them do violence to the fairly carefully designed texture of the presentation:

First paragraph Page 41· "insomnia" (sic !) is written for the Italian word "insomma" meaning "in short".

Third paragraph of the same page : "If you like" should have been the beginning of the sentence beginning, "He was studying how.. "

Fifth paragraph of the same page: "so, indeed Shakey had been provided.. " should have read, "so, indeed, long before Schank's cri-de-coeur about the impotence of AI programs, Shakey had been provided...."

Fourth paragraph of the same page: "moving a ramp" should have read "for example moving a ramp".

Yours sincerely,
Bernard Meltzer
Via G. Galilei 5,
210276 ISPRA (VA), Italy

The Magazine regrets having made any distortions of Prof. Meltzer's article. Although some errors were clearly egregious (e.g. changing "insomma" to "insomnia"), and should have been caught by us, some responsibility for the editing must be shared by our authors, who are given copies of the edited manuscripts for their approval prior to publication. — Ed

Update on the Autoling System

Editor:

William Katke's reference to the AUTOLING system in his paper, "Learning Language Using a Pattern Recognition Approach," (Spring, 1985, p. 69) is incomplete. AUTOLING was a true expert system, and represented an attempt to replicate the function of a human linguist interacting with a live informant. The conference presentation cited by Katke later appeared in a much longer work that included analyses of language examples in English, Latin, Roglai, Indonesian, Thai, and Mandarin Chinese (Klein *et al.*, 1968). Live, hands-on demonstrations via remote hookup to a Burroughs 5500 were given at the Linguistic Society of America Annual Meeting in 1967, and at colloquia at Carnegie-Mellon, UCLA and USC in 1968. A later version of the system was able to learn transformational grammars, and to guarantee, for a finite corpus, that it could parse all sentences previously recorded as valid, and, at the same time, fail to parse all sentences previously recorded as invalid (Klein & Kuppin, 1970). The method was extended to learning semantic network structures in combination with syntax, and a hand simulation of the analysis of a complex problem in Japanese that involved embedding constructions appeared in Klein (1973, 1978).

References

Katke, W (1985) Learning Language Using a Pattern Recognition Approach the *AI Magazine* 6:64-73

Klein, S (1973) Automatic Inference of Semantic Deep Structure Rules in Generative Semantic Grammars UWCS Tech Rep 180, Also in Computational and Mathematical Linguistics: Proceedings of the 1973 International Conference on Computational Linguistics, Pisa, A Zampoli (Ed), Florence: Olschki

Klein, S , W Fabens, R Herriot, W Katke, M A Kuppin & A Towster (1968) The AUTOLING System. UWCS Tech. Rep 43,

Klein, S & M A Kuppin (1970) An Interactive, Heuristic Program for Learning Transformational Grammars *Computer Studies in the Humanities & Verbal Behavior* 3: 144-162

Sheldon Klein
Department of Computer Sciences
University of Wisconsin
1210 West Dayton Street
Madison, Wisconsin 53706

Response to Franklin

Editor:

I assume that Stan Franklin's letter (AAAI profile, Vol 6 No. 1, Spring, 1985) can only be taken as a warning by the local arrangements committee of *IJCAI*: they had better do a better job of providing slim, white, wedding ring-less women for all those quiet, cool, athletic men who will be attending.

Sandra Eveland
International Business Machines Corporation
11400 Burnet Road
Austin, Texas 78758

(P.S. I can't believe you printed that letter.)

A Relativistic Approach

Editor:

The speed at which artificial intelligence is growing and the complexity and variety of the problems it is trying to solve raise some major issues of two different kinds:

1 The identity of AI (Where is AI going? Which are the goals of AI research now?)

2 The theoretical foundations and a general framework in which the different research fields of AI can fit (Which models are we using? What are they for? What are the theoretical implications of such models?)

Evidently, a change is taking place. At least two developments are possible:

1 A "freeze" in basic research (the "winter" of AI is a term that was used during the last AAAI Conference in Austin, August 1984)

2 A momentary loss of identity, seen as a temporary breakdown, while we look for a new AI research identity.

The second possibility, which is certainly more appealing than the first, implies quite a few changes in the actual way of thinking about AI research.

First, a substantial change is necessary in the way practical results and AI products are evaluated; it is evident that basic AI research needs some time to get a strong theoretical background, in order to produce better evaluation criteria for its own achievements and results.

The need for time and adequate research resources gets to be more and more significant: If we do not want to produce "canned research" for "ready to hand" products, we need to think of products as being just one and not the main aspect of research—even though real applications of a theory are the only way to test its validity.

In order to "create better criteria for evaluating AI research" and to get to applications too, everybody who is in AI today must be aware of the necessary gap between the current theoretical models and the actual products. If such awareness were missing, a whole set of problems and misunderstandings could deeply affect the way research in AI is evolving. If no autonomy were recognized to the theoretical part and if the theoretical part were to be reduced just to a set of *ad hoc* mini-theories to get "working programs" out of them, AI research could never achieve its primary goal, which is the setup of a theoretical background and of a general unifying framework where different AI subfields could fit together.

Every AI researcher knows that applications are a very important aspect too. How to get out from the apparent contradiction? By being relativistic and not reductivistic.

Being relativistic in AI means basically the following: Being aware of the gap between the current level of research and the need of the market that certainly influences the way research is being done. It also means that when AI researchers decide to create and use a partial model (of learning, of knowledge representation, of text-understanding, etc.,) they necessarily produce a set of evaluation criteria that are relative to those partial models that have been chosen and thus are not generally valid.

Being reductivistic in AI means the following: Being unaware of the limits of a partial model and considering such a model as being generally valid. This implies the reduction of a complex problem (in learning, in knowledge representation, in text-understanding, etc.) to a partial consideration and to the partial setup of a model. A reductivistic approach in AI research implies a natural consequence:

- Production of partial and not connected micro-models and microtheories, whose validity is more or less arbitrarily motivated
- The taking-for-granted that a partial modeling of the human mind can be justified and motivated in itself without needing any further testing or any psychological evidence
- The tendency of using other sciences (cognitive

sciences, psychology, linguistics, etc) as a support to give evidence to some very particular "*ad hoc* models" created such that you "just get a working program out of it "

Evidently, a reductivistic approach in AI research is simultaneously tempting and deleterious. It is tempting because it looks as if you were really getting somewhere and actualy achieving results; it is deleterious because it prevents you from continuing to do research by making you feel happy with partial results.

As research in AI proceeds, the real "potential" of the field progressively shows up. New problems arise, whose complexity shows the need of more and more "basic research." Since the field is growing so fast, it also becomes evident that partial models as well as partial theories are a good way to get started, but do not represent final results.

It also becomes evident that everyone in AI is necessarily working on a reduced model of intelligence; this is not a "bad thing" or a "good thing" in itself, but is just needs some more specifications. In other words, AI researchers are to be aware of the real potential and limits of their own models that is, every hypothesis has to be a part of a framework where answers must be given to the following set of questions:

- What is a certain model for?
- Which theory is it referring to, if any?
- What is this theory good for (aiming toward)?
- Has such theory been arbitrarily chosen (somehow motivated or supported by evidence and good for results)?
- Which are the aspects of such a theory that can be generalized and which are not?
- Which are the aspects of the problem that are not handled by the model?
- What relevance do they have to the general framework?
- Have other theories already been produced to handle such problems? How compatible are they with the present one?
- If there are problems, what are they and where do they arise?
- Are such conflicts generated by the use of differing and incompatible frameworks and evaluation criteria?

In conclusion, a relativistic approach does not mean discouraging research in AI by going back to some sort of "very weak feeling" about AI goals and achievements. Rather, it means being able to recognize partial development in AI by always keeping an eye on the "potential" of such research, without getting stuck at limited—even if relatively good—results

Graziella Tonfoni
Università di Bologna
Via Zamboni 16, 40126 Bologna, Italy

Letters to the editor should be addressed to Letters Editor, the AI Magazine 445 Burgess Drive, Menlo Park, California 94025, and should include the sender's address and telephone number. Not all letters can be used. Those that are will often be edited and excerpted.