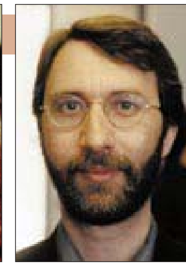


Marti A. Hearst
University of California, Berkeley
hearst@sims.berkeley.edu

Haym Hirsh
Rutgers University
hirsh@cs.rutgers.edu

AI's greatest trends and controversies

The transition to the next millennium gives us an opportunity to reflect on the past and project the future. In this spirit, we have asked a set of distinguished scholars and practitioners who were involved in AI's formative stages to describe, in just a few paragraphs, the most notable trend or controversy (or nontrend or noncontroversy) during AI's development.

The responses provide an interesting characterization of AI—and, in many ways, of the people of AI. We gave our contributors a great deal of flexibility in the nature of their responses. Some provided grand summaries of the history of the field as a whole. Others commented insightfully on more focused topics. Some observed changes and changed along with them. Others are still making advances on research agendas articulated presciently long ago. Some are optimistic. Others are pessimistic. Despite the range, both individually and collectively they provide insights into where we have been and where we are going.

Although each contribution is a unique expression of its author's glimpse back through AI's development, we are pleased to see the repetition of important themes that are at our discipline's core. Rather than simply presenting this as a list of contributor sound bites, we are instead weaving them into a narrative that showcases these themes as we go along. We hope this first Trends & Controversies of 2000 will serve as an interesting record of where AI is today, plus set the stage for what's to come.

—***Marti Hearst and Haym Hirsh***

Except perhaps for the AI naysayers, AI practitioners are creators—of software artifacts, their underlying algorithms, and their underlying theories. We begin our feature with Herbert Simon, one of our three contributors (together with John McCarthy and Oliver Selfridge) who are our links to the landmark Dartmouth conference in 1956, where modern AI is often said to have begun. Simon paints a broad picture of AI as a discipline constantly pursuing computational creations that challenge the uniqueness of biologically grounded intelligence.

Herbert A. Simon, Carnegie Mellon University

AI has been thought controversial because it challenged the uniqueness of human thought, as Darwin challenged the uniqueness of human origins. The boundaries of AI continue to expand rapidly, settling the controversy for those who know the evidence.

AI first demonstrated that important intellectual tasks could be accomplished by selective heuristic search, often in a thoroughly human way. GPS is one product of that line of research. Then AI explored the role of large bodies of knowledge in expert thinking. Dendral was an early important success, as was the extensive research on human chess expertise, modeled with such programs as Chrest (not Deep Blue, which is only partly humanoid). A third successful line has been the research on learning—for, example, Siklossy's ZBIE program, which learned natural language by comparing sentences with pictures.

Finally, there has been the great recent advance in robotics, based on progress in simulating sensory and motor functions.

The basic strategy of AI has always been to seek out progressively more complex human tasks and show how computers can do them, in humanoid ways or by brute force. With a half-century of steady progress, we have assembled a solid body of tested theory on the processes of human thinking and the ways to simulate and supplement them.

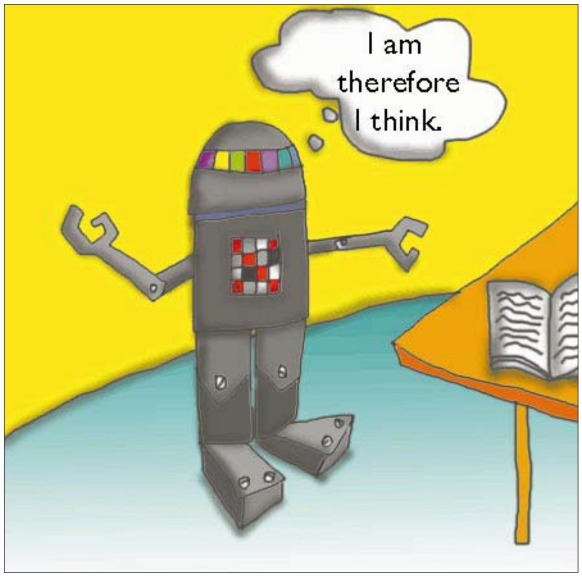
Knowledge representation and reasoning

In what media do AI practitioners create? The answer to this question is a depiction of AI itself, so it is not too surprising that most of our contributions address this question. Wolfgang Bibel, a proponent of the formalist agenda in AI, argues for the need for sophisticated logic formalisms and inferential methods for AI. Alan Bundy adds to these arguments, discussing further the advances achieved by those taking the formal-logic approach to AI, especially in light of the critiques raised by those on the other side of the AI fence.

Wolfgang Bibel, Darmstadt University of Technology

Among the controversies in AI, none is as persisting as the one about logic's role in AI. It can already be found in the early book *Computers and Thought* and is still around, as David Waltz's presidential address in the fall 1999 issue of *AI Magazine* demonstrates. On the one hand, there are those in AI (including me) who consider the language as well as the inferential machinery of logic as fundamental for the endeavor of realizing an artificial intelligence. Their (or at least my) short chain of arguments is as follows. Currently, natural intelligence manifests itself exclusively in terms of natural means of communication, foremost in natural language. Hence, intelligence would best be modeled in terms of natural language and its underlying mechanisms, which so far still is too hard to achieve. Logic is an abstracted, hence simpler form of natural language. Thus it provides an ideal medium for a first approximation.

On the other hand, this longstanding view has always been attacked with a variety of arguments. As far as I can tell, none of these arguments has been convincing enough to damage the logicist position in any substantial way—which does not mean that those attacks have not been successful in some other ways. For instance, we have been forced to live with many variations of logic. Also, we must as yet admit failure in two fundamental technical issues. One is the lack of a compilation of logical mechanisms rendering them efficient enough for practical applications (pursued, for example, in the area of program synthesis). The other concerns the integration of the great variety of specialized logical mechanisms within a single powerful, versatile, and yet uniform system. This only demonstrates that the logical way is a hard one and takes more time than initially expected, which happens to be true for AI as a whole.



AI after Y2K

Alan Bundy, University of Edinburgh

I entered the field of AI in 1971 when I joined Bernard Meltzer's group at the University of Edinburgh. This small group specialized in automated theorem proving and contained a number of people who went on to become academic stars: Bob Kowalski, Pat Hayes, Bob Boyer, and J. Moore. Most work in ATP at that time was based on refinements of resolution—for example, Kowalski's SL resolution, which became the basis of Prolog.

The major controversy at that time was the criticism of so-called uniform proof procedures, which mainly came from MIT, such as by Minsky, Papert, and Sussman. Complete and general provers, such as those based on resolution, were seen as being incapable of proving non-trivial theorems or of application to commonsense reasoning because of the combinatorial explosion. More domain-specific alternatives were being explored. For instance, Hewitt's Planner programming language was designed with various built-in control facilities under user control. These criticisms significantly reduced the efforts in automated theorem proving. It inspired a number of developments in specialized, effective, but incomplete inference mechanisms in AI, which continue to this day.

Meanwhile, a handful of researchers around the world continued to pursue resolution theorem proving. Efficient indexing and storage schemes were developed, and the increasing power of modern computers was exploited. This work has confounded some of the conventional wisdom of the '70s. A string of open conjectures have been solved, especially by the group of Wos, McCune, and others at Argonne National Labs. The most famous of these was the settling of the Robbins conjecture, which had been open since the '30s. Its solution by machine made newspaper front pages across the world.

Although Drew McDermott still shares many of Bibel's and Bundy's formalist leanings, his view has changed significantly over the years. The one-time proponent of the pure formal-logic AI agenda articulates here his vision, where AI's success arises not from sophisticated representation and reasoning methods, but rather from simple representations and tractable algorithms, where complexity is mastered in the process of formulating a problem so that such methods can be applied. Hans Berliner takes this even one step further—if successful AI bottoms out at simple methods such as brute-force search, then one of the most important contributions to AI has been Moore's Law and the relentless increases in computer speeds.

Drew McDermott, Yale University

In my opinion, the most important trend is the decline in complex knowledge representations and the success of propositional or probabilistic competitors. By complex representations, I mean those that attempted to include quantification over variables, inheritance, structural descriptions of objects, and diverse inference schemes. These systems have attracted great interest because of an essentially *semantic* argument, of this form: People can have thought *H*; formal system *S* is capable of representing *H*; therefore *S* is interesting. Unfortunately, these arguments haven't often gotten down to computational brass tacks. The theories tend to float in the ether, mainly to be used as inspiration.

Meanwhile, the successful computational theories have either had no representations at all (sometimes called procedural representations) or have been based on general-purpose engines for manipulating notations such as Bayes nets or propositional logic. These are general-purpose in the sense that they don't know that the formulas they manipulate have to do with planning or natural language. However, it often requires great ingenuity to translate a new problem into the notation, because it usually lacks all the handy features (such as variables) that make notations flexible. Instead, a symbol stands for a single proposition that has a single number representing the degree to which it is believed (0, 1, or a probability). What you get for the pain of fitting into this straitjacket is that your problem actually gets solved.

I expect this trend to continue and for the original semantic argument to turn out to be a mirage. It will turn out that everything worth doing can be done with simple representations.

Hans Berliner, Carnegie Mellon University

I consider the most important trend was that computers got considerably faster in these last 50 years. In the process, we found that many things for which we had at best anthropomorphic solutions, which in many cases failed to capture the real gist of a human's method, could be done by more brute-forcish methods that merely enumerated until a satisfactory solution was found. If this be heresy, so be it.

Ed Feigenbaum's contribution goes one step higher and argues for the need for knowledge representation, countering recent efforts that claim that the world in which an intelligent artifact operates is the suitable and necessary representation for building successful AI artifacts. Feigenbaum then turns the focus around from knowledge representation to knowledge itself, arguing that when AI has had success in complex reasoning tasks, it has typically depended on encoding and exploiting knowledge of the domain. Feigenbaum also provides an important admonition to critics of AI to be constructive rather than opaque in their cynicism toward AI. Finally, Bruce Buchanan shares Feigenbaum's focus on the importance of knowledge, but paints a middle ground. The long-standing clashes between formal logic-based approaches and knowledge-based techniques lay not in disagreement about which approach is right but rather from agreement about the importance of commonsense reasoning for AI, with differences centering only on where the common sense should lie.

Edward A. Feigenbaum, Stanford University

There is no question that the "mainstream" path of development of AI since 1966 has been the path of knowledge-based systems for achieving intelligent behavior in programs. The first decade of AI research concentrated on process, particularly the problem-solving processes (such as GPS or theorem proving) that were called, by Newell and Simon, weak methods. But these methods became strong only by adding substantial bodies of domain-specific knowledge. The proof of this came with the proliferation of powerful expert systems (knowledge models of expertise). Yet AI researchers moved from process to content (knowledge) only very slowly because knowledge is domain-specific—hence, viewed as someone else's field of work, and smacked of application, not CS research. Minsky made this his theme in his Turing Award lecture in the late 1960s, entitled "Form vs. Content in Computer Science." Yet the issue remains strong to this day.

In the 1980s, a backlash to the mainstream knowledge-based approach developed. Call it *anti*representation or *0*representation. This was an antiknowledge representation (antimodeling) view. In this view, comprehensive knowledge modeling of the real world was much too difficult; therefore, the real world itself had to serve as its own best model. The role of the cognitive agent was to respond effectively and adaptively to actual situations encountered in interactions with the real world. Much experimentation has shown that relatively simple behaviors can be achieved in this way, but there are no examples of complex cognitive behavior (such as expert-level problem solving in a domain) that has been achieved in this manner.

Both of these are themes and controversies within AI. But the idea of AI has always touched raw nerves in some circles outside AI. Some philosophers have objected, with arguments that are sometimes vague, sometimes elaborate, that "something else" is involved in thinking that is not modeled by information-processing programs for computing machines. AI researchers have countered that, although it might take decades to work out all the scientific details, there is nothing else other than information processing underlying thought. Other skeptics are practitioners in subdisciplines of computer science other than AI. They simply cannot believe that the relatively simple computers on which they write their database programs, operating systems, graphics, or numerical algorithms can be the substrate for models of thought. Critics have offered nothing but their skepticism and negativism. Science thrives on criticism, providing that the critics offer something else as a counter to something. The controversy can then be resolved by further experiment and theory making. But the critics of the AI enterprise have been particularly unhelpful in this regard.

Bruce G. Buchanan, University of Pittsburgh

The most substantive controversy in AI that I participated in was framed, at the time, as whether intelligence derived primarily from logical reasoning or from knowledge. It was thought by many, unfortunately, to be an exclusive *or*.

The Dendral program was an existence proof that computers could couple technical knowledge with simple inference mechanisms—no more sophisticated than *modus ponens*—to reproduce the results of highly intelligent scientists.[1] Ed Feigenbaum went on record early with the "knowledge is power" theme. I was struggling with engineering the knowledge of chemists into production rules so that we could use one simple inference mechanism and extend its scope solely by writing new declarative statements, conditional rules, and specialized predicates. It was the model for Mycin and the whole first generation of knowledge-based systems.[2,3]

The logicists in AI at that time were building on the success of resolution theorem proving and looking for extensions to first-order logic that would encompass commonsense modes of inference. Default reasoning will always be important in some form—for example, when specialized knowledge has been exhausted or none of it seems to apply. So the controversy was never about whether common sense is necessary but how much it should be built into the logic or into the knowledge base.[4] Evidence to date shows that successful applications can be built as knowledge-based systems while we are waiting for the final answer.

1. R.K. Lindsay et al., *Applications of Artificial Intelligence for Chemical Inference: The DENDRAL Project*, McGraw-Hill, New York, 1980.
2. B.G. Buchanan and E.H. Shortliffe, *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, Mass., 1984.
3. E.A. Feigenbaum and B.G. Buchanan, "DENDRAL and Meta-DENDRAL: Roots of Knowledge Systems and Expert System Applications," *Artificial Intelligence*, Vol. 59, Nos. 1–2, 1993, pp. 223–240.
4. D.B. Lenat and E.A. Feigenbaum, "On the Thresholds of Knowledge," *Artificial Intelligence,* Vol. 47, Nos. 1–3, 1991, pp. 185–250.

Machine learning

One tantalizing cognitive task that has been central to AI since its earliest days is learning. Both Oliver Selfridge and Donald Michie identify learning as the key to AI's future. Although Selfridge grounds his arguments on the primacy of learning in human behavior, Michie focuses on the recent con-

vergence of advances in both learning and deductive reasoning. Both lend strong credentials to their arguments—Selfridge's dating back more than 40 years, including his participation in the 1956 Dartmouth conference, and Michie's dating back to his even earlier code-breaking work with Alan Turing during World War II.

Oliver Selfridge, MIT and GTE/BBN

"Can you think of some chore or duty that a person does that she doesn't do better the second time? Or can you think of some chore or duty that a computer does that it does do better the second time?"

A chief essence of human intelligence is continuing learning or adapting, and we rely on that all the time. It is change and the handling of change that triggers learning and adapting: we do not program our children with rigid rules, because they will have to change anyhow, through learning and adapting. Machine learning is the most important aspect of AI, and not just learning particular skills or subroutines, but rather, everything, including behavior, cognition, symbolic manipulation, models of the world, and, above all, goals and what to want. An especially profound piece of learning is called common sense, to which increasing attention is being paid.

This suggests that AI software should be more concerned with being changeable—and all that that implies—than with satisfying specifications; that is, with being correct. Machine learning is growing steadily, and in that field I include both neural nets and genetic programming, because both are occupied with dealing with changes. In people, though, learning and adapting take place at many levels simultaneously and continually, and we in machine learning must find out how to build our models and software to do so too.

"Find a bug in a program, and fix it, and the program will work today. Show the program how to find and fix a bug, and the program will work forever."

Donald Michie, Artificial Intelligence Applications Institute, University of Edinburgh

Reviewing the 50 years since Alan Turing's 1950 paper in *Mind*, I see the most notable trend in AI as springing from the recent confluence of two subtrends, each in itself notable—namely,

- machine learning, experimentally initiated by A.L. Samuel in the mid 1950s and lifted to concept-learning level by Earl Hunt in the early 1960s, and
- automation of deductive reasoning, formulated as a real-world problem by John McCarthy in the late 1950s and underpinned by J.A. Robinson's resolution algorithm in the mid-1960s.

The subsequent confluence, today termed inductive logic programming, was initiated by Banerji and Sammut in the early 1980s and independently during the same period by Ehud Shapiro, after an earlier theoretical insight by Gordon Plotkin. Automated induction has since been further developed by Stephen Muggleton and others and applied to theory formation in a number of applied sciences. The reason for attributing high AI importance flows from the nature of scientific discovery, widely regarded as one of the more conspicuous and consequential manifestations of articulate human intelligence.

The most notable nontrend has resulted from consistent disregard of the closing section, Learning Machines, of Turing's 1950 paper. A two-stage approach is there proposed:

1. Construct a teachable machine,
2. Subject it to a course of education.

Far from incorporating Turing's incremental principle, even the most intelligent of today's knowledge-acquisition systems forget almost everything that they ever learned every time their AI masters turn to the next small corner of this large world. As someone who knew Alan Turing well and what he had in mind, I rate this as the most significant nontrend by far of the half-century. Let us now change this.

AI: An interdisciplinary challenge

What will it take to achieve AI? Three of our contributors argue for the need for broad literacy well beyond the confines of any single discipline. Nils Nilsson points to the breadth of disciplines he has had to study over the course of his long involvement in AI. Despite seeing intelligence as a complex and interdisciplinary phenomenon, he is nevertheless optimistic about the enterprise. Aaron Sloman's complementary contribution points to the slow, if steady, progress that AI has been making and the range of complex problems that still challenge us as demonstration of the indisciplinary breadth of challenges that we face. David Waltz discusses how the complexity of many human cognitive tasks was a surprise to many, echoing the arguments of Nilsson and Sloman about the need for broad literacy in AI. Waltz also points out the intriguing possibility that the development of fMRI might soon enable a bottom-up agenda for AI based on study of the human brain.

Nils Nilsson, Stanford University

AI shows all the signs of being in what the late Thomas Kuhn called a pre-paradigmatic, pre-normal-science stage. It has many ardent investigators, arrayed in several camps, each claiming to have the essential approach to creating intelligence in machines. We have logicists, Bayesians, connectionists, evolutionists, intensive computationalists, heuristic searchers, control theorists, reinforcement learners, reactivists, embodied computationalists, and on and on. It might be that *intelligence* (of the sort that we would want in flexible, autonomous robots, for example) is the kind of multiplex for which no single science or paradigm will ever emerge. Each of these approaches, and possibly more, will contribute its part to an overall design.

For my part, I tell students that to be successful in future AI research they will have to be familiar with several disciplines. Among those that I have had to learn a bit about are statistics, machine learning, logic, control theory, linguistics, computer science, discrete and continuous (!) mathematics, cognitive psychology, philosophy of mind, and ethology, as well as all of the core AI techniques that have been explored over the last 50 years or so. Yes, people will make some specialized advances from a base of just one or two of these, but to achieve human-level AI (which has always been my goal) we'll need Renaissance people who can draw on a wide variety of disciplines. I hope some are out there!

How to Reach Us

Writers
For detailed information on submitting articles, write for our Editorial Guidelines (isystems@computer.org), or access computer.org/intelligent/edguide.htm.

Letters to the Editor
Send letters to

Managing Editor
IEEE Intelligent Systems
10662 Los Vaqueros Circle
Los Alamitos, CA 90720
dprice@computer.org

Please provide an e-mail address or daytime phone number with your letter.

On the Web
Access computer.org/intelligent for information about *IEEE Intelligent Systems.*

Subscription Change of Address
Send change-of-address requests for magazine subscriptions to address.change@ieee. org. Be sure to specify *Intelligent Systems.*

Membership Change of Address
Send change-of-address requests for the membership directory to directory.updates@computer.org.

Missing or Damaged Copies
If you are missing an issue or you received a damaged copy, contact membership@computer.org.

Reprints of Articles
For price information or to order reprints, send e-mail to isystems@computer.org or fax +1 714 821 4010.

Reprint Permission
To obtain permission to reprint an article, contact William Hagen, IEEE Copyrights and Trademarks Manager, at whagen@ieee.org.

Intelligent SYSTEMS
& their applications

Aaron Sloman, University of Birmingham

AI has two themes:

- engineering (making useful things) and
- science (investigating natural intelligence and what sorts of intelligence are possible, and how).

AI as engineering has had many small but useful advances, often hailed as major revolutions by their promoters (one notable trend!).

AI as science moves very slowly, revealing

- what the problems are and
- why all the plausible mechanisms are inadequate.

AI as science is multidisciplinary: it feeds on and contributes to neuroscience, psychology, linguistics, logic, biology, social sciences, computer science, software engineering, mathematics, and philosophy.

Progress is very slow. We understand little about the functions of vision in human beings. We can see far more than physical properties of objects. We see functions, causal relations, happiness in a face, gestures, and what Gibson called affordances. But what all that means is not clear. Because we can't yet accurately characterize visual abilities of squirrels, or humans, we are nowhere near explaining or replicating them.

An often-noted related human capability is spatial reasoning. But nobody has a good characterization of what that means. So all proposed models are far from adequate.

And then there's consciousness and emotions—much hyped and little understood.

Progress might come from an important recent trend, away from focusing on forms of representation and algorithms toward considering types of architectures for complete systems. This might lead to important new developments, especially if instead of looking only at what normal intelligent adults can do we also consider infants, people with various kinds of brain damage, and other kinds of animals, providing clues to our evolutionary history, shedding new light on what we are now and how we might be modeled and replicated. We need to explore our neighborhood in design space. But we also need to attend to fine-grained results, from neuroscience and so forth.

David Waltz, NEC

The most striking trend from my perspective has been the steady increase in my estimate of the complexity of duplicating human mental performance. When I was a grad student, I believed that modeling introspection and language would be sufficient. Experimental psychology was likely to be useful to AI, but we would not need to understand neuroscience, and emotions and clinical psychology were irrelevant. (It was hard to see how we'd ever be able to ethically measure what was going on inside human brains in any detail.) Logic would need to be modeled, and it would need to be extended to handle representations of events. Computations were likely to be expensive—there was no notion of NP-completeness—so heuristics would be needed. Even as late as 1980, the expert systems community still held beliefs fairly similar to these, although the need for more complex logics for general reasoning had become evident.

Today, it is hard to see how we would have missed the vast complexities required to model generic human-like thought, let alone the thought of a particular individual. There are many different reasoning styles, heavily influenced by individuals' experiences (such as using analogical reasoning from precedents, not just logic). The problem of generating an event or situation representation from sensory inputs is wildly underspecified, and dependent on current individual goals. Also, fMRI has made it likely that neuroscience will play a major role in informing AI (for example, distributed representations seem well established).

The big question remaining is whether distance to the biggest goals of AI will continue to remain constant or increase as time goes on, or whether we're closing the gap.

Other contributors have taken up some of the difficult issues discussed above. Rod Brooks points out that how perception works is still a mystery and must be understood in order to model intelligence. The joint contribution by Randall Davis and Howard Shrobe expands on the need for further study of the use of visual representations in AI. Margaret Boden discusses the complexity in achieving such distinctly human traits as emotion and creativity. Ryszard Michalski, on the other hand, articulates the challenge of building intelligent systems that help people capitalize on the overwhelming range of information that modern society is constantly producing.

Rodney Brooks, MIT

For me, the most important change in AI happened in the 1980s when some people realized that the model of reasoning used in AI was very different from what happens inside the heads of people, very different at any level of abstraction used for the descriptions. Such differences do not invalidate the nonhuman approaches—airplanes are good examples of very useful machines that operate very differently from the way real birds operate. But the realization in AI opened up new ways of doing things and new avenues to go down, especially in the area of robots embedded in the real physical world. This, coupled with the vast increase in onboard computer power that we can put on robots, has let us push against the boundaries of what our artificial perceptual systems can do. And this leads us to the current conundrum in which we find ourselves.

The mass of the human brain is perhaps 50% devoted to perception. We now have enough computer power to rival the neural systems of insects and small animals, but we do not know how to organize that computation to do perception, or at least generate the dynamics of perception coupled to action, anywhere near as well as those simple creatures do. We have gotten pretty good at motion vision from a stable platform and at face detection in limited circumstances, but are not very good at motion vision from a moving platform, general face detection (at the level at which a sheep can perform, say), facial feature understanding, facial recognition, general vision for navigation, and object segmentation, and we are truly hopeless at object recognition (a subject on which I and countless others have written PhD theses) in general circumstances.

We know that in biological systems adaption or learning is used extensively in all these areas, but we have not seen good ways to connect any current understanding

Coming Up Next

Does KDD Uncover New Insights?

from machine learning to these domains. We await a few good maverick young students to have some critical insights and show us how blind we old fogies have been for the last few years.

Randall Davis and Howard Shrobe, MIT

For much of its life, AI has been heavily influenced by what it saw in its early stages as exemplars of intelligence—namely, the sort of symbol processing found in books such as *Principia Mathematica* and in tasks such as playing chess, proving theorems, or planning, all of which involve reasoning with textual representations. We've made substantial progress on these tasks but have paid relatively less attention to reasoning with other modalities of representation.

Examples of using other modalities include the work done on reasoning with diagrams and the work in robotics on embedded intelligence, both of which develop representation and reasoning methods distinct from the traditional textual approach. Still, the notion of diagrams as a basis for intelligent inference, visual imagery, and related notions have received insufficient attention, it seems to us, especially considering how prevalent they appear to be in our everyday intelligence and in our brain anatomy. It would be useful and possibly important to take seriously the notion of diagrams and images as fundamental representations for intelligence.

Another way to come at this is to say that perhaps vision (and sensory modalities) ought not to be treated as input channels whose information must be condensed and turned into textual symbolic representations before intelligent processing can occur. Instead, there might be significant intelligence going on much earlier—when the information is still in the form of images. This is hardly a new speculation; the point is simply that relatively little work has taken it seriously, considering how plausible it is and how valuable the payoff it might provide.

Margaret A. Boden, University of Sussex

In the early days of AI, a high proportion of AI models dealt with motivation, emotion, personality, and social interaction. By the 1970s, this was no longer so. The problems were too complex to be fruitfully approached. Because they all involve the scheduling and prioritizing of several goals, they could not be properly addressed before AI research could cope with single-goal systems. Now, there is some promising research on computational architectures for motivation and emotion, on (still fairly primitive) implemented models of emotional intelligence, and on social behaviors such as cooperation and communication. I expect these areas to be increasingly important in the next century.

Another area of growing importance is AI creativity. Virtually no one in the early days mentioned creativity as an AI topic. But since then some important work has been done, on science, the arts, and everyday thinking (analogy). At present, the most intractable problem is how to enable the computer to evaluate its own novel "ideas." Until AI systems can be fruitfully (although not infallibly) creative, their ability to model—and even to aid—human thinking will be strictly limited.

The study of self-organization will be another crucial area. This topic has been addressed for many years in certain types of connectionism, and is now featured in various new ways in A-Life.

Finally, AI scientists must learn how to integrate the insights, and computational strengths, of different methodologies. Hybrid systems, much more ambitious than those known today, will be needed to model real intelligence.

Ryszard Michalski, George Mason University

By "intelligence," we mean a set of capabilities that let a system with limited resources (energy, time, and memory) operate under limited input information (incomplete, uncertain, inconsistent, or incorrect). To have such capabilities, a system must satisfy three conditions. It must

- have sensors that allow it to collect information;
- be able to turn information into knowledge (abstracted, organized, and generalized information)—that is, to be able to learn; and
- be able to use this knowledge to achieve its goals (inherited, embodied, or generated internally)—that is, to reason.

A system satisfying these three conditions is called intelligent. AI's goal is to understand and build intelligent systems.

One of the most important directions of AI (or computational intelligence) for the new century is to build intelligent systems that will let people access and utilize enormous amounts of information generated by modern societies in various forms (texts, images, speech, and multimedia) for making informed decisions. Such systems should be able to help people to derive target-oriented knowledge from information obtainable from diverse sources (the Internet, books, journals, and other people).

The practice of AI

All disciplines—AI included—have a social structure as well as an intellectual one. We also received contributions commenting on the practice of AI research itself. Jerry Feldman discusses the conflict between methodology-driven and phenomenon-driven research in AI, as well as the conflict between quantitatively evaluable research and the more creative and often less quantifiable research necessary for a constantly growing discipline. A second part of the contribution of Davis and Shrobe (which we chose to include here separately) also addresses this concern, focusing on its ramifi-

cations in acceptance standards in our discipline's journals and conferences. Hubert Dreyfus, a longtime critic of AI, also discusses what makes a field a science and emphasizes the need for self-criticism.

Jerry Feldman, University of California, Berkeley, and the International Computer Science Institute

The continuing conflict that has been most important for me is that between methodology-driven and phenomenon-driven research. AI defines itself by methodology: there are still many who equate AI with formalization in predicate calculus, and newer guilds are based on reinforcement, subsumption architecture, graphical models, and so on. There are powerful forces driving this social behavior, but it does conflict with the original AI goals of understanding and implementing intelligence. There was, for me, an exciting ecumenical period around 1980, but then connectionism was narrowed to cover only what could be studied using back-propagation in uniform layered networks.

What has been more disturbing is the almost total lack of conflict over the demise of basic research in AI (and most other fields as well). The standard explanation, which seems right to me, is that the end of the Cold War removed the motive of fear that was behind societal support of basic science. One can hope that the current requirement of immediate payoff from every effort will pass, but there is no sign of this. There is a federal law, the Government Performance and Results Act of 1993, that mandates numerical measures of progress for every funded project. Some of us continue to indulge in curiosity-driven research, but progress on the basic questions of intelligence is much slower than it could be with even modest societal support.

Randall Davis and Howard Shrobe, MIT

As a young science, AI has gone through predictable struggles over what ought to constitute a publishable result. Some early conference papers describe programs that were little more than one-trick ponies; others simply contained speculations about how a program might be built. Over time, more and more came to be demanded for publication, both in terms of a more substantial body of test cases and an increasing formality in characterizing results. The appropriate balancing point between innovative ideas and formally characterizable results has always been controversial. Even so, there is a feeling that the pendulum might have swung too far of late, demanding even of conference papers the sorts of development and maturity of ideas that are required of archival journal publications. This is problematic: If even conference papers require fully developed results, where is the venue for the untried, the innovative, the outlandish? Every field needs some degree of unconventional thinking to keep it fresh; there needs to be a place for this in AI that provides broad exposure. Conference papers, with their inherently limited length and limited lifetime, are an appropriate place and an appropriate venue to get the ideas into circulation in the community. Program committees should take this into account, with an appropriate set of standards for this category of papers.

Hubert L. Dreyfus, University of California, Berkeley

A mature science progresses by setting forth clear predictions and subjecting them to falsification, then asking why the prediction failed, which assumptions turned out to be sound, and which unjustified, and so learning from mistakes. By such standards AI is only half a science. In the first 20 years—from roughly 1960 to 1980—it had clear goals and forthright predictions as to how and when they would be achieved. The goal was programming computers to perform in intelligent ways, and the method was to use symbolic knowledge representations.

That symbolic AI clearly failed to reach its goals is to its credit. Being able to fail means the claims were not vacuous, and recognizing failure leads to a science progressing by forming new hypotheses or undergoing a revolution. What always amazes me is that no one in AI ever stops to contemplate his or her setbacks and disappointments. They either claim, like Simon, that what they were setting out to do, they did, or, like Lenat, are about to do, or, like McCarthy, is harder to do than they realized, but no one seems to ask whether they were perhaps on the wrong track.

Rather, they simply show they were on the wrong track by abandoning some approaches—for example, rule-based expert systems and large commonsense knowledge databases—and turning toward modest, sober, engineering problems and statistical approaches. It's striking that we read nothing about ambitious projects like CYC and COG: neither that they have failed nor that they are making progress. It's as if physicists at the turn of the previous century had hushed up their failure to find the ether drift and just turned away from cosmology to more modest projects. There is nothing wrong with such a pragmatic approach. But a lack of self-criticism does make a mockery of the claim that AI is a mature and slowly progressing science, and it exposes AI to the risk, noted by Santayana, that those who don't understand their history are doomed to repeat it.

The big picture

We near our journey's conclusion with several sweeping views of the discipline's history. First, Roger Schank focuses on AI's quest to model and understand the human mind. He touches upon themes broached by many of our other contributors—and includes irreverent pokes in a number of directions, such as AI's entrepreneurial '80s and those with formalist proclivites. Saul Amarel, on the other hand, discusses the field's maturity in terms of its transition to work grounded in real-world problems. Amarel, alone among our contributors, identifies the (ongoing) vital role that government funding agencies have had in the progress in our discipline. Barbara Hayes-Roth reflects on the many strands of research that make up our field, emphasizing that this rich mosaic of efforts—and often controversies—is what collectively defines our enterprise.

Roger C. Schank, Northwestern University

Once upon a time, AI was about the mind and how it worked. The main controversy of the old days was whether computers should attempt to model how things were done by people or whether computers should follow methods that no person could ever do and that might actually improve on people. Those who followed the first path considered themselves to be cognitive scientists. Those who followed the latter path were either engineers (anything that works is good) or formalists (mathematical elegance and the pretense of science were exciting to those folks).

The engineers had their day in AI for a while. They made promises of what could be done with rule-based systems that all the cognitive scientists knew could never be achieved. Unfortunately, the venture capitalists didn't know this, and AI went through a period of rapid expansion and just as rapid decline that still causes people in the business world to shudder at the mention of AI. The formalists seem to be in control these days, looking for more mathematics to put in their papers and trying ever harder to make it seem like AI is a science.

Cognitive science is still alive and well, but AI and cognitive science don't have much to say to each other any more. What I learned in attempting to model mind on a computer is that learning is the most interesting thing people can do and that to imitate real learning we must start at the beginning. We cannot just stuff a computer full of facts that someone else has learned. People learn by doing and so must computers. Computers will have to have things they can do, the ability to process and reflect upon what they have done, and a world with which to interact, in order to become sentient beings.

Saul Amarel, Rutgers University

The late '50s and '60s were AI's heroic age. Most of the key ideas and approaches in the field came up at that time. There was excitement in the air, and every new development was greeted with much enthusiasm by the few researchers (mostly in the US) who worked in the field. ARPA support, motivated primarily by the goal of developing computer-based approaches to complex decision-making processes in defense, was essential for getting AI research off the ground and for providing the continuity of support needed to build solid research teams in the area. Work on AI methods grew typically in the context of relatively simple tasks that were under complete control of the researcher, such as puzzles, games, and artificial environments. This was the age of discovery, invention, and constructive exploration, and there was a sense of being at the cutting edge of computer science.

The '70s saw ambitious attempts to bring AI methods to problems of interpretation and diagnosis in science, medicine, and defense. Large bodies of domain knowledge were needed to solve such problems, and various issues of knowledge handling started to attract researchers' attention. The field of knowledge-based expert systems grew up in this period. It became increasingly apparent that the performance of AI systems depends on careful synergy with other computing assets—machine architectures, languages, databases, and interfaces. Building a system that works well in the real world proved to require sizable amounts of time and effort. At the same time, some basic problems of representation, search, and reasoning that started to receive attention in the previous decade proved to be more difficult than originally expected. As the decade progressed, there was a growing criticism of AI in several quarters—both about its basic ideas and methods and about its (potential for) practical impact.

An important event of the '80s was the launching of the strategic computing program at ARPA. A central goal of the program was to bring AI out of the Lab, and to explore its power in solving complex real-world problems with the support of advanced high-performance computing and software technologies. This program contributed to substantial progress in AI and to a better integration of AI with other parts of the computing field. Interestingly, the major beneficiary of strategic computing was high-performance computing, which became the centerpiece of the HPCC national initiative that followed the strategic computing program.

At present, AI has reached a reasonable level of scientific maturity as part of computer science, and it looks more like a conventional discipline. There is steady incremental progress in the discipline. In several parts of the field, the emphasis is on refinement rather than on development of new ideas or exploration of new territory. I believe that the field will benefit enormously from the exposure to several demanding real-world challenges (such as problems in science and engineering and in management) in the context of several large initiatives where AI researchers can stress the boundaries of existing methods and explore new ideas as they interact with researchers from other disciplines. The present surge of national interest in IS/IT research provides a promising environment for such a development to take place.

Barbara Hayes-Roth, Stanford University

In my opinion, the most interesting and persistent controversy in AI concerns the relationship between artificial intelligence and human intelligence. Like the bird to the flying machine, human intelligence was the founding inspiration for artificial intelligence. Almost immediately, there ensued a series of intense debates about whether we could, should, or even want to build machines that mimic human intelligence. Human intelligence has been viewed and then rejected as the manifestation of a particular mechanism to be replicated in silicon. But also like the bird, it has been viewed and then actively pursued as a functional demonstration of the behaviors to be performed by some to-be-determined alternative mechanism.

Some AI scientists meticulously study and recreate the performance properties of human intelligence, believing that "God is in the details." Other scientists, disdaining the human intelligence "airplane," claim to aim straight for the rocket of superhuman intelligence. Some researchers decompose human intelligence into complementary components and then focus on some of these in isolation from the others. Other researchers argue that only an integrated approach to the complete system can succeed. Some AI scientists see both human and artificial intelligence as the triumph of pure reason, whether that be logic, mathematics, statistics, or economics, while others build innovative application systems whose power derives from simple operations performed on expert or everyday knowledge.

Although there are many differences of opinion and even heated arguments in this multithreaded debate, there is also a continuing movement forward in our understanding of and capabilities for artificial intelligence. The only position that I find discouraging is the premature conclusion of impossibility. Some reach this conclusion based on the observation that we have not yet achieved a spectacular success. But how can we give up after such a small effort in our young field? Others conclude impossibility based on philosophical sophistry, which might just as well persuade us not to await the sunrise tomorrow.

Researchers who do not believe in the possibility of a human-like artificial intelligence need not pursue it. In the meantime, we should encourage the proliferation of diverse perspectives and approaches and the continuing controversy that give our field both life and promise as a long-term scientific enterprise and practical value as a near-term practical enterprise.

Final thoughts

We would have liked to include many more people in this special Trends & Controversies. In some cases, the lack of a contribution by your favorite AI authority is due to space constraints imposing a modest-sized list of people we could contact and to a very positive response from the members of that list. In other cases, individuals were on our list but politely declined. In two cases, our request inspired initially cynical reactions, which we include here. Judea Pearl explains how controversies, although apparently wasting time and energy, can have value in forcing work in directions in which it would not ordinarily go. We conclude with the words of John McCarthy, the man who gave the discipline of AI its name. To paraphrase: Enough talk, get back to work!

Judea Pearl, UCLA

I was about to write that controversies are a waste of time. Indeed, each time I open my book (*Probabilistic Reasoning*), I can kick myself for having spent so many pages on defending probability calculus vis-à-vis its then competitors (such as rule-based systems, fuzzy sets, belief functions, or even logic), and I can't help speculating on where we would be today had all this energy been spent on more constructive, nonapologetic research.

However, as soon as I keyed in those words, I recalled some specific incidents from the '80s, which lead me toward the opposite conclusion: It was in response to challenges from competing paradigms that some latent capabilities of probability theory were uncovered in those years. For example, belief propagation in Bayes networks was developed to compete with the modularity of rule-based systems. Similar motivations stood behind developments in nonmonotonic reasoning, iterative belief revision, and theories of actions and causation. In summary, the virtues of controversies from my observations have been to jolt complacent experts into uncovering dormant capabilities in their areas of expertise.

Sold on the virtues of controversies, I now find myself alarmed that my latest thoughts on counterfactuals (*Causality*, forthcoming) have evoked no controversy at all (at least in AI)—I hope someone takes this note as a challenge.

John McCarthy, Stanford University

Commenting on trends and controversies is not a high priority with me. I'd rather work on AI. Even taking part in a controversy would be better.