The Promises and Problems of Artificial Intelligence

By Leah A. Lievrouw
Department of Information Studies
& Communication Studies Program, UCLA

Lately I've heard some interesting comments about artificial intelligence (AI). At a day-long conference on media design theory at a Los Angeles-area design school last spring, one attendee blithely remarked that AI has made computers able to "think for us." More recently, over lunch, a colleague opined that new AI speech recognition techniques, coupled with wireless technology, would soon overcome the digital divide.

Though often derided as an oxymoron, the term "artificial intelligence" was coined in 1956 to describe hypothetical computers that could understand and solve problems as humans do, that is, machines that could "think." Computers were seen as analogous to "brains" or "minds" (Edwards, 1996). Depending on one's point of view, since then AI has either been a brilliantly innovative generator of most of today's "smart" technologies and intelligent machines, or the most overhyped, publicly funded money pit of postwar R&D.

Advocates point to speech and vision recognition systems, computer graphics, "interactive" media, the Internet and sophisticated information retrieval systems as evidence for AI's success. Critics (including AI researchers themselves) say that despite years of work computers remain incapable of even the simplest types of "learning" or general-purpose problem solving envisioned by the founders of the field. A decade ago, one columnist remarked, "AI has yielded some very useful technologies and products. It just hasn't produced any true artificial intelligence" (Davis, 1990, p. 162). The eminent AI scientist Marvin Minsky said more recently,

> Around 1980, progress [in AI] stopped in some ways . . . it stopped because we'd done the easy things. . . . As far as I know, nobody has been able to get a machine to solve real problems that are informally expressed, the way somebody would normally express them.

What accounts for these widely divergent accounts of AI's accomplishments and significance? Why does it inspire such uncritical confidence, on one hand, and such disillusionment, even cynicism, on the other? Recent books and commentaries provide some insights into a field, now nearing 50 years old, that has become part of U.S. industry and popular culture but whose promise seems unfulfilled.

A basic history of AI is sketched by Paul Edwards in *The Closed World* (Edwards, 1996). He places it within the "interdiscipline" of cognitive science, which has developed "a view of the human as an information processing system, in a precise, quantifiable sense . . . under three kinds of influences: military needs, contact with communications engineers, and experience with communications hardware and concepts" (p. 212). Like its midcentury intellectual forebears (cybernetics, cognitive psychology, operations and systems research, and Claude Shannon's theories of information and communication), Edwards argues that AI was shaped by two discourses: the "closed world" of Cold War/post-1989 American military power and political influence, and the cyborg or "machine in the middle" discourse applied to the design and deployment of postwar weapons and communications systems.

What distinguishes AI from the other areas of cognitive science, however, is its attempt to blur the line between biological and machine intelligence. If cybernetics sought to model *brains* in computer hardware, and the project of cognitive psychology was to model human *minds* or intelligence as cybernetic machines, AI attempted to "mimic minds in software" (Edwards, 1996, p. 239). AI shifted from biological to symbolic models of knowledge and intelligence to become "thinking about thinking," "[presenting its] . . . programs as psychological theories" (Dreyfus, 1979, p. 18).

Symbolic processing was the conceptual centerpiece of AI through its first 30 years. AI programs represented "knowledge" as fairly restricted and rule-bound sets of objects and relationships. However, by the early 1970s, supposedly intelligent programs had proven incapable of "understanding" everyday language and context. In a *New York Times* interview in 1998, Marvin Minsky recalled a program in the 1960s that was supposed to read and solve problems from a high school algebra book, but failed because it could not "read" the language the problems were written in. Subsequently, scientists built a program that could read a first-grade children's story: "We could build into the program the knowledge necessary to read that story. We didn't have much trouble with the grammar." However, the program failed "as soon as something was mentioned that the program didn't know about . . ." (Dreifus, 1998, p. F3).

In 1972 computer scientist and philosopher Hubert Dreyfus published a scathing and polemical challenge to AI, *What Computers Can't Do* (Dreyfus, 1972). He not only criticized the inability of AI programs to model taken-for-granted contextual knowledge; he expressed serious doubts that machine intelligence could ever be realized so long as "knowledge" was conceptualized as symbolic processing:

> . . . intelligence requires understanding, and understanding requires giving a computer the background of common sense that adult human beings share by virtue of having bodies, interacting skillfully with the material world, and being trained into a culture. (p. 3)

He wasn't alone in his critique. In the 1970s researchers began to limit their claims and restrict their experimental projects to artificially constrained "microworlds" (such as the tabletop "blocks world" of Terry Winograd's SHRDLU program), or to highly structured, rule-bound domains of expertise, such as the derivation of chemical structures (Edward Feigenbaum's DENDRAL) or medical diagnosis and drug therapy (Edward Shortliffe's MYCIN). The latter "knowledge engineering" approach extracted and represented the specialist's knowledge to build "expert systems." Within these rigidly defined domains, the programs did seem fairly robust.

However, in the late 1970s Dreyfus extended his critique to these newer programs because they were not generalizable beyond the specific domains for which they were written (Dreyfus, 1979). Furthermore, he said, the implicit assumption that microworlds and expert systems could eventually be combined to create a broad representation of the "world" or "knowledge" was inappropriate for modeling human knowledge and experience because it was based on a reductionist view of knowledge borrowed from the natural sciences.

AI researchers took the contextual challenge seriously as the failures began to add up. "Almost everyone now . . . agrees that representing and organizing commonsense knowledge is incredibly difficult, and that facing up to this problem constitutes the moment of truth for AI" (Dreyfus, 1979, p. 3). Terry Winograd, among others, recognized that the microworlds approach he had helped to create had played out and that a different way was needed to bring context into knowledge representation. For example, with collaborator Fernando Flores, Winograd incorporated speech act theory and principles from Heideggerian language philosophy to build The Coordinator, a program to support collaborative work and decision making. Winograd and Flores's approach was fundamentally communicative, but not in the information-processing sense: "Communication is not a process of transmitting information or symbols, but one of commitment and interpretation" (Winograd & Flores, 1987, p. 176).

More recently, my UCLA colleague Phil Agre has proposed a similar move for AI from what he calls the "mentalism" of cognitive science to an "interactionism" that models action and engagement with people and things in the everyday world (Agre, 1997). Among other AI researchers, the symbolic processing school of thought has given way to theories that seem closer to the "machine brain" metaphor of early cybernetics. These include connectionism, parallel processing, and even the neural network models originally developed by Warren McCulloch and Walter Pitts in the 1940s (Edwards, 1996, p. 188). With the benefit of newer supercomputer hardware, these machine/software combinations seem better able to "learn" from context, rather containing all predefined relevant knowledge in the system from the outset.

Though the story seems to focus on the failures rather than the successes of AI, this thumbnail history illustrates its intellectual and philosophical maturation and the sophisticated level of critique within the field. But what about the confidence and power that so many people outside the field seem to attribute to AI?

It may well be, as Davis (1990, p. 162) said skeptically, that "having faith in AI is somewhat akin to believing in world peace: Both are wonderful dreams, but we'll probably never live to see either." On the other hand, perhaps over the last half-century "the meaning of AI has changed" (Coffee, 1998, p. 48). Instead of producing a general theory of intelligence, its main contribution has been smaller-scale techniques and products, adapted and incorporated into other technologies: "The AI comet has shed some useful chunks of stuff" (Coffee, 1998, p. 48). Dreyfus himself noticed the shift in the late 1970s:

> In recent years the attempt to produce special-purpose programs tailored to narrowly-restricted domains . . . has been abandoned by AI *theorists* and frankly and quite successfully taken over by self-styled AI *engineers*, with no interest in making generally intelligent machines. (emphasis in the original; Dreyfus, 1979, p. 3)

In the 1960s and 1970s, search and optimization techniques "spun off" as core components of early database and information retrieval systems. In the 1980s expert systems techniques were applied in private industry. AI algorithms, memory manipulation techniques, pattern matching and rule/classification systems helped manage large, distributed data collections. Today, speech recognition and cheap vision systems seem to be the most promising spinoffs.

So we might conclude that both perspectives are right: AI techniques are a pervasive part of the new media infrastructure, though machine intelligence on a par with human understanding remains speculative. In its next 50 years, new directions in AI research might bring it closer to what the founders had in mind.

References

Agre, P. (1997). *Computation and human experience*. Cambridge, UK: Cambridge University Press.

Coffee, P. (1998, August 17). The Halley's comet of computing. *PC Week*, 15(33), p. 48.

Davis, F. (1990, December 3). Though the myth lives on, true AI is still just a dream. *PC Week*, 7(48), p. 162.

Dreifus, C. (1998, July 28). Why isn't artificial intelligence more like the real thing? A conversation with Dr. Marvin Minsky. *New York Times*, p. F3.

Dreyfus, H. L. (1972). *What computers can't do: A critique of artificial reason*. New York: Harper & Row.

Dreyfus, H. L. (1979). *What computers can't do: The limits of artificial intelligence* (rev. ed.). New York: Harper Colophon.

Edwards, P. N. (1996). *The closed world: Computers and the politics of discourse in Cold War America*. Cambridge, MA: MIT Press.

Winograd, T., & Flores, F. (1987). *Understanding computers and cognition*. Reading, MA: Addison Wesley (originally published in 1986 by Ablex, Norwood, NJ).