

## Evolution, Sociobiology, and the Future of Artificial Intelligence

David L. Waltz, *Columbia University*

**W**hat will AI systems be like in the near and long terms? Basically, we'll get the AI that people are willing to pay for. Consequently, many specialized applications will appear long before AI demonstrates its "Manifest

Destiny" of human-level general intelligence. The AI demonstrations and applications we're going to see in the near future will trend strongly toward "cognitive prostheses"—systems that do well things that humans do poorly or don't like to do. Both near-term and far-future systems will need to interact smoothly with humans, which will put special constraints on them. In particular, to build systems that we'll trust and want to use, we'll need to carefully consider and craft their implicit and explicit values.

### Near term

Over the next 20 years, we'll see the effects of evolutionary pressures on AI subareas. Some subareas will flourish, others will shrivel, some will die. The kinds of AI we'll be doing in 20 years will be determined by interactions between three factors:

- financial factors that dictate the subareas that receive funding;
- technical factors, especially genuinely useful applications we develop; and
- scientific factors, such as the areas where we achieve the greatest intellectual progress.

There will be virtuous and vicious cycles: valuable applications and scientific breakthroughs will attract funding, which will in turn drive expansion of those areas, and uninteresting and unprofitable areas will decline, despite some funding inertia and lobbying.

I'm sure that evolution will be at work on AI. We're moving out of AI's infancy, a period when all subareas—practical and theoretical—competed on a relatively equal footing, primarily by touting their potential impact. Given that all these subareas were primarily offering promises,

it's no surprise that scientifically interesting work has tended to dominate more mundane applications-directed engineering work. However, AI is moving into its young adulthood, with some subareas paying their own way. More resources are being steered into the areas that succeed financially, both for particular applications and for the modules and algorithms responsible for their success.

### Long term

In the long range, we'll see a kind of sociobiological competition among intelligent entities, resulting in the wide replication of particular autonomous systems and intelligent applications. The big winners will be systems that succeed in being "selected" by human users and funders: they'll be both economically valuable and socially congenial. Unless systems are congenial—dependable, pleasant to work with, and not irritating—they're unlikely to be widely propagated.

Ironically, such factors were foreseen in AI's prehistory in Asimov's three laws of robotics and in the Turing Test. The Turing Test really tests indistinguishability between a system and human in a social interaction, not a system's measured intelligence (such as the result of an IQ or SAT test).

### Evolutionary AI families and their prospects

AI today can be usefully grouped around five basic families, each with distinctive evolutionary prospects:

- *Expert AI.* The goal here is to build systems, possibly with superhuman abilities, for specialized applications. Examples typically require emphasizing particular aspects of intelligence and knowledge, not broad or general intelligence. The Conference on Innovative Applications of Artificial Intelligence showcases such work. Expert systems and chess programs are prime examples.
- *Autonomous robots.* The most ambitious version of this goal would be Turing Test AI plus perception, learning, and action. More probable goals are particular classes

of robots, each specialized for some task—such as mining, soldiering, or exploration—but not necessarily endowed with full, broad, human-level intelligence.

- **Cognitive prostheses.** The goal here is to build human-computer systems where computers do the things that they do best (such as analyzing large amounts of data) or that people don't like to do (such as dangerous or tedious tasks) in tight conjunction with people doing what they do best (such as general reasoning and planning). Combined systems will be super-human. This goal shades into human-computer interaction (HCI) systems and generally includes intelligent interfaces. Augmentation can be physical (for example, exoskeletons or remote sensors) as well as mental.
- **AI theory and algorithms.** This covers algorithms and specialized methods that come from AI but have taken on a life independent of their initial inspiration. This type of research is sometimes justified by its potential for advancing one of the other four goals listed here, but often the justification is implicit. A wide range of examples exists, including constraint programming, inductive-logic programming, search and planning algorithms, reasoning under uncertainty, and machine learning and pattern recognition algorithms.
- **Turing Test AI,** including both the actual Turing Test and its usual misinterpretations. This goal requires human-level intelligence, including language and reasoning, although not necessarily including perception and action in the strict Turing Test. Two variants are possible here: a *cognitive science Turing Test AI*, which would try to map human mental organization, and a *software-engineered Turing Test AI*, which would attempt to mimic behavior but not necessarily model the details of human cognition.

Of course, these five goals don't represent crisply differentiated categories; much R&D combines two or more types of goals (for example, systems in DARPA's Grand Challenge desert race overlap both the AI expert and autonomous-robot areas without fully belonging to either). Still, for the purposes of mapping the evolutionary future, we can usefully understand AI as centering on one or another of these targets.

### Expert AI

Autonomous systems with expertise in particular areas are greatly important and will be big winners over the next 20 years. The prime opportunities today are for applications of machine learning and data mining technologies that can produce high levels of expertise with relatively little specific programming. The first wave of expert systems floundered, largely because building such systems was labor intensive. It also required artistry on the part of the "knowledge engineer" to appropriately represent knowledge in a form that general reasoners could use. Machine learning has provided expert AI with orchards of "low-hanging fruit"—applications with relatively high payoff and low risk. Early machine learning methods (such as backpropagation neural networks, genetic algorithms, Bayesian nets,

Along with expert AI, the cognitive prostheses area is the most likely of all AI goals to continue to flourish, because it offers many opportunities for near-term advances and low-hanging fruit.

and case-based reasoners) required a lot of manual effort to put data into a form suitable for the learning algorithms. Today's most widely used machine learning algorithms (such as support vector machines, boosting, and genetic programming) require minimal data preparation and can deal with high-dimensional data, requiring only labeled training sets or explicit representations of goal states.

### Autonomous robots

By adding sensors and effectors, even a narrow or subhuman intelligent system can have real value. Such systems will require us to combine several kinds of abilities in a coordinated cognitive architecture, a recent hot area at DARPA and NASA and for industries (consider Sony and Honda's humanoid robots). This research area has many potential applications, the most prominent of

which are battlefield and rescue robots (for reconnaissance, bomb detection and removal, and so on), as well as caretaker and companion robots, space-exploration robots, and self-driving vehicles. All these can eventually become great businesses, and their important military and domestic-security applications will likely insure continuing government funding and rapid progress. Learning gives a huge advantage in this area as well—Sebastian Thrun's Stanley was able to win the DARPA Grand Challenge less than a year after the project was begun, largely because Thrun was able to exploit learning methods to construct controls for many of its subsystems.

### Cognitive prostheses

Along with expert AI, this general area is the most likely of all AI goals to continue to flourish because it offers many opportunities for near-term advances and low-hanging fruit. Semantic Web applications are especially likely to drive progress in this area in the near term. Funding should be relatively plentiful. This area overlaps with psychology and ergonomics and will also benefit from progress in those areas.

### AI theory and algorithms

This area represents the wild card—the subarea where new ideas can yield the greatest upside surprises. New algorithms can work magic—for example, turning lengthy theoretical demonstrations into real-time applications or replacing human engineering with learning. It's crucial to continue to support research in these areas, but this kind of work is the most threatened because it can't realistically promise practical results. Today in the US, only the National Science Foundation funds this kind of research directly, and success rates for NSF theory grants are very low, in the 10 percent range. Of course, grants with more practical overall goals, both from the NSF and other agencies, typically also include at least some research in theory. Industrial funding in theory is, and will likely remain, minimal.

### Turing Test AI

Of these five goals, Turing Test AI is perhaps the goal that people outside AI most commonly ascribe to AI researchers. At the same time, Turing Test AI is the goal least likely for AI researchers to actually pursue. The commonly misunderstood version of the Turing Test—building a system indistin-

guishable from a human in a conversation—is commonly seen as AI’s “Manifest Destiny.” Nonetheless, there’s no reason to believe that human intelligence is the natural limit for any intelligent system. Indeed, if a system equal to a human is built, natural advances in hardware speed will eventually lead to a system that performs at superhuman speed and will thus constitute a superhuman intelligence (just as a system that reasons at human levels but takes much longer to come to its conclusions wouldn’t be considered as smart as a human).

Overall, the prospects for such a system are quite poor. First, few agencies or industries are likely to fund research whose primary goal is to pass some variant of the Turing Test. Why would anyone want such a system? Human beings are plentiful, and hiring one is quite inexpensive compared to the foreseeable R&D, capital, and expense costs of a computer system that could pass the test. Moreover, without sensors and effectors or expert knowledge, such a system would have minimal application value, except perhaps as a companion or conversational partner.

### Who will fund future AI research?

In the kind of world I suggest we’re entering, will anyone be prepared to fund the pursuit of a full-fledged system capable of broad, human-level intelligent reasoning, perception, language, and learning? I suspect not. If this goal is to be realized, it will be a side effect of research in the other areas, so progress will be slower than if it were a direct goal.

Some basic research will be done at companies, but most basic work in these areas will likely be done in academia. Looking back over history, only near-monopolies (pre-breakup AT&T Bell Labs, IBM, and Xerox PARC up through the early ’90s, Microsoft and possibly Google and Yahoo today) have funded basic industrial AI research over extended periods.

Nonetheless, AI has reached the stage where it can pay its way in building new applications, where many power tools (such as software packages for learning, planning, or processing natural language or images) are available, data is plentiful, and hardware advances have made it possible to broadly field fast, affordable systems centered on or including AI. Both corporate and university research targeted toward

applications will benefit. We can expect more products with AI content. It’s not so clear that there will be many AI start-up companies—AI will likely be a component, not a complete application, and much R&D work will be development with at most a small, basic research contribution. Industrial funding of university applications-oriented research is a distinct possibility and one that my lab at Columbia University has been pursuing, especially with Con Edison, New York City’s power supplier.

The shrinking of DARPA and NASA budgets at the theory end and shortages of NSF money threaten US basic research in AI theory and algorithms. DARPA has moved toward large, joint university-commercial integrator projects requiring fundamental advances but also featuring two or three yearly competitions that weed out relatively

Will anyone be prepared to fund the pursuit of a full-fledged system capable of broad, human-level intelligent reasoning, perception, language, and learning? I suspect not.

weak teams. Good ideas alone aren’t sufficient in such an environment. Only teams including large universities with professional staffs can expect to compete successfully. NASA has cut most research that’s not directly related to building a new space shuttle. It’s unlikely that either industries or other countries will pick up the slack in theory; other countries will likely mirror US trends. To the extent that basic theoretical advances will be necessary to fully realize the more applied AI goals, progress toward these goals will slow down. We will come to regret low funding in these areas, but the people responsible for current policies will be long gone from their positions before the emptying of the research pipeline becomes evident.

On the other hand, theory and algorithms work is relatively inexpensive (compared, say, to autonomous robotics research) and

progress will still occur, albeit more slowly than we’ve been used to. I believe generally in the “Society of Minds” view—that intelligence is the result of many systems operating on different principles, a consequence of which is that no single algorithmic advance will suddenly make intelligence possible. However, a potent learning algorithm that could recapitulate brain evolution isn’t inconceivable and could change everything.

Academically, AI has gone from being a separate part of computer science curricula to one that has had its results absorbed into other specialties—for example, learning into security systems, natural-language processing into HCI, image and video processing into graphics, and learning methods into databases.

AI R&D will continue to be a healthy activity, but we can expect it to result in smart, adaptive aids to humans; narrowly capable autonomous robotic systems; and superhuman experts. Turing Test AI isn’t likely any time soon.

### AI individuals as the endpoint of AI’s evolution

AI seems to generally assume that the yardstick for measuring human-level intelligence is clear—we’ll recognize it when we see it. But as with the human genome—the first versions of which were actually the genomes of particular persons—as many human intelligences exist as people, living and dead. Moreover, unlike the human genome, which is all nature, each human intelligence has a strong nurture component. Turing cleverly picked for his test a practically implementable measure—the ability to successfully convince a questioner of one’s gender, measured statistically—leaving aside any attempts to directly measure intelligence.

AI has generally taken the attitude that intelligent action will be provably logically correct or optimal in some sense. Supposedly, an intelligent system will try to optimize whatever task it’s pursuing—in games the system will try to win, in diagnosing disease it will try to be correct, and so on. Humans, who can be masochists, ascetics, clowns, or suicide risks, among other irrational stances, exhibit no single shared set of values. Apparently self-defeating attitudes might have subtle social or sexual goals (for example, trying to get sympathy or bring out another person’s parental tendencies). So, critically, systems that inter-

act with irrational humans will require not just rationality but also a deep understanding of human nature, including human biological imperatives. Systems likely won't be able to experience directly the deepest human motivations: competition and survival via food, clothing, shelter, bonding, reproduction, parenting, and so on. For example, to fool a Turing Test questioner who is trying to determine gender, the system must be able to simulate and understand a range of human needs and desires.

Cute robots—Kismet, Aibo, Qrio, and Asimo come to mind—tap into human social tendencies, and attention to social interactions is a recognized part of HCI research. But will systems seem so cute or appealing when vast numbers of identical clones of each type exist? Or will we need to give systems personalities so that we're not driven mad by repetitious and totally

predictable interactions with them?

Our intelligent systems will need to learn about us and our motivations in all our variety. We might want our systems, at least those that we deal with in nonemergency situations, to be driven by internal social goals—for example, to be accepted as colleagues by humans who interact with them or to be considered interesting, surprising, or amusing. In short, to be judged truly intelligent, systems might need personalities.

**A**I's progress, especially in machine learning technology, has brought us to a point where we can offer valuable, practical systems and modules. This will likely drive AI's evolution in a way that distorts the field's shape, greatly enlarging funding and activities at the applications and engineering



**David L. Waltz** is the director of the Center for Computational Learning Systems at Columbia University. Contact him at [waltz@ccls.columbia.edu](mailto:waltz@ccls.columbia.edu).

end of the spectrum, possibly at the expense of the scientific and big-picture end.

Nonetheless, the spread of AI will be limited to useful applications unless and until we can build systems that people bond with and want to have as part of their lives. Once we can build such systems—which must necessarily be able to understand human motivations, needs, and tastes—people will readily spread AI. Broadly intelligent autonomous systems will be the descendants of such useful, congenial systems. ■

**PURPOSE** The IEEE Computer Society is the world's largest association of computing professionals, and is the leading provider of technical information in the field.

**MEMBERSHIP** Members receive the monthly magazine *Computer*, discounts, and opportunities to serve (all activities are led by volunteer members). Membership is open to all IEEE members, affiliate society members, and others interested in the computer field.

#### COMPUTER SOCIETY WEB SITE

The IEEE Computer Society's Web site, at [www.computer.org](http://www.computer.org), offers information and samples from the society's publications and conferences, as well as a broad range of information about technical committees, standards, student activities, and more.

#### BOARD OF GOVERNORS

**Term Expiring 2006:** Mark Christensen, Alan Clements, Robert Colwell, Annie Combelles, Ann Q. Gates, Robit Kapur, Bill N. Schilit

**Term Expiring 2007:** Jean M. Bacon, George V. Cybenko, Antonio Doria, Richard A. Kemmerer, Itaru Mimura, Brian M. O'Connell, Christina M. Schober

**Term Expiring 2008:** Richard H. Eckhouse, James D. Isaak, James W. Moore, Gary McGraw, Robert H. Sloan, Makoto Takizawa, Stephanie M. White

**Next Board Meeting:** 16 June 06, San Juan, PR

#### IEEE OFFICERS

**President :** MICHAEL R. LIGHTNER

**President-Elect:** LEAH H. JAMIESON

**Past President:** W. CLEON ANDERSON

**Executive Director:** JEFFREY W. RAYNES

**Secretary:** J. ROBERTO DE MARCA

**Treasurer:** JOSEPH V. LILLIE

**VP, Educational Activities:** MOSHE KAM

**VP, Pub. Services & Products:** SAIFUR RAHMAN

**VP, Regional Activities:** PEDRO RAY

**President, Standards Assoc:** DONALD N. HEIRMAN

**VP, Technical Activities:** CELIA DESMOND

**IEEE Division V Director:** OSCAR N. GARCIA

**IEEE Division VIII Director:** STEPHEN L. DIAMOND

**President, IEEE-USA:** RALPH W. WYNDRUM, JR.

IEEE  
computer  
society  
60<sup>TH</sup> anniversary

#### COMPUTER SOCIETY OFFICES

##### Washington Office

1730 Massachusetts Ave. NW

Washington, DC 20036-1992

Phone: +1 202 371 0101

Fax: +1 202 728 9614

E-mail: [bq.ofc@computer.org](mailto:bq.ofc@computer.org)

##### Los Alamitos Office

10662 Los Vaqueros Cir., PO Box 3014

Los Alamitos, CA 90720-1314

Phone: +1 714 821 8380

E-mail: [belp@computer.org](mailto:belp@computer.org)

##### Membership and Publication Orders:

Phone: +1 800 272 6657

Fax: +1 714 821 4641

E-mail: [belp@computer.org](mailto:belp@computer.org)

##### Asia/Pacific Office

Watanabe Building

1-4-2 Minami-Aoyama, Minato-ku

Tokyo 107-0062, Japan

Phone: +81 3 3408 3118

Fax: +81 3 3408 3553

E-mail: [tokyo.ofc@computer.org](mailto:tokyo.ofc@computer.org)



#### EXECUTIVE COMMITTEE

##### President:

DEBORAH M. COOPER\*

PO Box 8822

Reston, VA 20195

Phone: +1 703 716 1164

Fax: +1 703 716 1159

[d.cooper@computer.org](mailto:d.cooper@computer.org)

**President-Elect:** MICHAEL R. WILLIAMS\*

**Past President:** GERALD L. ENGEL\*

**VP, Conferences and Tutorials:**

RANGACHAR KASTURI (1ST VP)\*

**VP, Standards Activities:** SUSAN K. (KATHY) LAND (2ND VP)\*

**VP, Chapters Activities:**

CHRISTINA M. SCHOBERT\*

**VP, Educational Activities:** MURALI VARANASI†

**VP, Electronic Products and Services:**

SOREL REISMAN†

**VP, Publications:** JON G. ROKNET†

**VP, Technical Activities:** STEPHANIE M. WHITE\*

**Secretary:** ANN Q. GATES\*

**Treasurer:** STEPHEN B. SEIDMAN†

**2006–2007 IEEE Division V Director:**

OSCAR N. GARCIA†

**2005–2006 IEEE Division VIII Director:**

STEPHEN L. DIAMOND†

**2006 IEEE Division VIII Director-Elect:**

THOMAS W. WILLIAMS†

**Computer Editor in Chief:** DORIS L. CARVER†

**Executive Director:** DAVID W. HENNAGE†

\* voting member of the Board of Governors

† nonvoting member of the Board of Governors

#### EXECUTIVE STAFF

**Executive Director:** DAVID W. HENNAGE

**Assoc. Executive Director:** ANNE MARIE KELLY

**Publisher:** ANGELA BURGESS

**Associate Publisher:** DICK PRICE

**Director, Administration:** VIOLET S. DOAN

**Director, Information Technology & Services:**

ROBERT CARE

**Director, Business & Product Development:**

PETER TURNER

rev. 6 March 06