If Not Turing's Test, Then What?

Paul R. Cohen

■ If it is true that good problems produce good science, then it will be worthwhile to identify good problems, and even more worthwhile to discover the attributes that make them good problems. This discovery process is necessarily empirical, so we examine several challenge problems, beginning with Turing's famous test, and more than a dozen attributes that challenge problems might have. We are led to a contrast between research strategies—the successful "divide and conquer" strategy and the promising but largely untested "developmental" strategy—and we conclude that good challenge problems encourage the latter strategy.

Turing's Test: The First Challenge

More than fifty years ago, Alan Turing proposed a clever test of the proposition that machines can think (Turing 1950). He wanted the proposition to be an empirical, one and he particularly wanted to avoid haggling over what it means for anything to think.

> We now ask the question, 'What will happen when a machine takes the part of [the man] in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"

More recently, the test has taken slightly different forms. Most contemporary versions ask simply whether the interrogator can be fooled into identifying the machine as human, not necessarily a man or a woman.

There are many published arguments about Turing's paper, and I want to look at three *kinds* of argument. One kind says Turing's test is irrelevant; another concerns the philosophy of machines that think; the third is methodological.

Ignore It, and Maybe It Will Go Away…

Blay Whitby (1996) offers this humorous history of the Turing test:

> 1950–1966: A source of inspiration to all concerned with AI.

> 1966–1973: A distraction from some more promising avenues of AI research.

> 1973–1990: By now a source of distraction mainly to philosophers, rather than AI workers.

> 1990: Consigned to history.

Perhaps Whitby is right, and Turing's test should be forgotten as quickly as possible and should not be taught in schools. Plenty of people have tried to get rid of it. They argue that the test is methodologically flawed and is based in bad philosophy, that it exposes cultural biases and naïveté about what Turing calls the "programming" required to pass the test. Yet the test still stands as a grand challenge for artificial intelligence, it is part of how we define ourselves as a field, it won't go away, and, if it did, what would take its place?

Turing's test is not irrelevant, though its role has changed over the years. Robert French's (2000) history of the test treats it as an indicator of attitudes toward AI. French notes that among AI researchers, the question is no longer, "What should we do to pass the test?" but, "Why can't we pass it?" This shift in attitudes—from hubris to a gnawing worry that AI is on the wrong track—is accompanied by another, which, paradoxically, requires even more encompassing and challenging tests. The test is too behavioral—the critics say—too oriented to language, too symbolic, not grounded in the physical world, and so on. We needn't go into the details of these arguments to see that Turing's test continues to influence the debate on what AI can or should do.

There is only one sense in which Turing's test is irrelevant: almost nobody thinks we should devote any effort in the foreseeable future to trying to pass it. In every other sense, as a historical challenge, a long-term goal for AI, a philosophical problem, a methodological case study, and an indicator of attitudes in AI, the Turing test remains relevant.

Turing the Philosopher

Would Turing mind very much that his test no longer has the role he intended? If we take Turing at his word, then it is not clear that he ever intended his test to be attempted:

> There are already a number of digital computers in working order, and it may be asked, 'Why not try the experiment straight away?...' The short answer is that we are not asking whether all digital computers would do well in the game nor whether the computers at present available would do well, but whether there are imaginable computers which would do well.

Daniel Dennett thinks Turing intended the test as "a conversational show-stopper," yet the philosophical debate over Turing's test is ironically complicated. As Dennett says, "Alas, philosophers—amateur and professional—have instead taken Turing's proposal as a pretext for just the sort of definitional haggling and interminable arguing about imaginary counterexamples he was hoping to squelch" (Dennett 1998).

Philosophers wouldn't be interested if Turing hadn't been talking about *intentional* attributes of machines—beliefs, goals, states of knowledge, and so on—and because we in AI are about building machines with intentional attributes, philosophers will always have something to say about what we do. However, even if the preponderance of philosophical opinion was that machines can't think, it probably wouldn't affect the work we do. Who among us would stop doing AI if someone proved that machines can't think? I would like to know whether there is life elsewhere in the universe; I think the question is important, but it doesn't affect my work, and neither does the question of whether machines can think. Consequently, at least in this article, I am unconcerned with philosophical arguments about whether machines can think.

Turing's Test as Methodology

Instead I will focus on a different, entirely methodological question: *Which attributes of tests for the intentional capabilities of machines lead to more capable machines?* I am confident that if we pose the right sorts of challenges, then we will make good progress in AI. This article is really about what makes challenges

I am confident that if we pose the right sorts of challenges, then we will make good progress in AI.

good, in the sense of helping AI researchers make progress. Turing's test has some of these good attributes, as well as some really bad ones.

The one thing everyone likes about the Turing test is its *proxy function,* the idea that the test is a proxy for a great many, wide-ranging intellectual capabilities. Dennett puts it this way:

> "Nothing could possibly pass the Turing test by winning the imitation game without being able to perform indefinitely many other intelligent actions. ... [Turing's] test was so severe, he thought, that nothing that could pass it fair and square would disappoint us in other quarters." (Dennett 1998)

No one in AI claims to be able to cover such a wide range of human intellectual capabilities. We don't say, for instance, "Nothing could possibly perform well on the UCI machine learning test problems without being able to perform indefinitely many other intelligent actions." Nor do we think word sense disambiguation, obstacle avoidance, image segmentation, expert systems, or beating the world chess champion are proxies for indefinitely many other intelligent actions, as Turing's test is. It is valuable to be reminded of the breadth of human intellect, especially as our field fractures into subdisciplines, and I suppose one methodological contribution of Turing's test is to remind us to aim for broad, not narrow competence. However, many find it easier and more productive to specialize, and, even though we all know about Turing's test and many of us consider it a worthy goal, it isn't enough to encourage us to develop broad, general AI systems.

So in a way, the Turing test is impotent: It has not convinced AI researchers to try to pass it. Paradoxically, although the proxy function is the test's most attractive feature, it puts the cookie jar on a shelf so high that nobody reaches for it. Indeed, as Pat Hayes and Ken Ford point out, "The Turing Test is now taken to be simply a rather fancy way of stating that the goal of AI is to make an artificial human being" (Hayes and Ford 1995).

A second notable methodological failing of Turing's test is that it pushes many aspects of intelligence into one test that has a yes or no answer. This isn't necessary. We could follow the lead of the multiple intelligences movement in cognitive psychology and devise tests of different sorts of intelligence. In fact, Turing's test is not even very complete, when viewed in terms of, say, Howard Gardner's catalog of intelligences (Gardner 1983). It focused mostly on logical, linguistic, and interpersonal intelligence, not on intrapersonal, bodily-kinesthetic, naturalist, musical, and visual-spa-

tial intelligence (rounding out the eight in Gardner's catalog).

Robert French goes further and criticizes the test for its focus on culturally oriented human intelligence: "The Test provides a guarantee not of intelligence but of culturally-oriented human intelligence" (French 2000). The test also says nothing about neonatal or infant intelligence—which I think are worth understanding and emulating. In fact, to the extent that Turing provided an operational definition of intelligence with his test, it was a kind of middle-class English intellectual, midcentury dinner party kind of intelligence and represents neither the pinnacle nor the broad plain of human intellectual capability.

A third failing of the test is that it cannot be passed today. Tests should be challenging, but tests that *cannot* be passed provide no information. Periodically the Loebner Prize organization conducts a sort of Turing test. Consider the following brief snippet of a dialogue I had with one of the entrants:

PC: Where is your nose?

Entrant: Oh, my nose? Why are you so interested?

PC: Is it near your eyes?

Entrant: Such is Human Perversity.

PC: Is your nose between your mouth and your eyes?

Entrant: I can't tell you that.

PC: How many hands do I have?

Entrant: I don't understand why you have, and furthermore, I don't really care.

PC: My daughter's name is Allegra. What is my daughter's name?

Entrant: Is she a big part of your problem?

It takes only a moment to see I am not conversing with a human being. My point isn't to make fun of the people who try to win the Loebner Prize, nor do I think this snippet is the best that we can do with today's technology. My point is that even the very best technology in AI today would not bring us anywhere close to passing the Turing test, and this has a very bad consequence: Few AI researchers try to pass the test.

Said more positively, a good test is only slightly out of reach, and the path to success is at least partly clear.

Not only is Turing's goal remote, but attempts to pass his test are not diagnostic: They don't tell us what to do to pass the test next time. Blay Whitby puts it this way: "If the Turing test is read as something like an operational definition of intelligence, then two very important defects of such a test must be considered. First, it is all or nothing: it gives no indication as to what a partial success might look like. Second, it gives no direct indications as to how success might be achieved" (Whitby 1996). And Dennett notes the asymmetry of the test: "Failure on the Turing test does not predict failure on ... others, but success would surely predict success" (Dennett 1998). Attempting the test is a bit like failing a job interview: Were my qualifications suspect? Was it something I said? Was my shirt too garish? All I have is a rejection letter—the same content-free letter that all but one other candidate got—and I have no idea how to improve my chances next time.

So let's recognize the Turing test for what it is: A goal, not a test. Tests are diagnostic, and specific, and predictive, and Turing's test is neither of the first two and arguably isn't predictive, either. Turing's test is not a challenge like going to the moon, because one can see how to get to the moon and one can test progress at every step along the way. The main functions of Turing's test are these: To substitute tests of *behavior* for squabbles about definitions of intelligence, and to remind us of the enormous breadth of human intellect. The first point is accepted by pretty much everyone in the AI community, the second seems not to withstand the social and academic pressure to specialize.

So now we must move on to other tests, which, I hope, have fewer methodological flaws; tests that work for us.

New Challenges

Two disclaimers: First, artificial intelligence and computer science do not lack challenge problems, nor do we lack the imagination to provide new ones. This section is primarily about *attributes* of challenge problems, not about the problems, themselves. Second, assertions about the utility or goodness of particular attributes are merely conjectures and are subject to empirical review. Now I will describe four problems that illustrate conjectured good attributes of challenge problems.

Challenge 1: Robot Soccer

Invented by Alan Mackworth in the early 1990s to challenge the simplifying assumptions of good old-fashioned AI (Mackworth 1993), robot soccer is now a worldwide movement. No other AI activity has involved so many people at universities, corporations, primary and secondary schools, and members of the public.

What makes robot soccer a good challenge problem? Clearly the problem itself is exciting, the competitions are wild, and students stay up

So let's recognize the Turing test for what it is: A goal, not a test.

late working on their hardware and software. Much of the success of the robot soccer movement is due to wise early decisions and continuing good management. The community has a clear and easily stated fifty-year goal: to beat the human world champion soccer team. Each year, the community elects a steering committee to moderate debate on how to modify the rules and tasks and league structure for the coming year's competition. It is the responsibility of this committee to steer the community toward its ultimate goal in manageable steps. The bar is raised each year, but never too high; for instance, this year there will be no special lighting over the soccer pitches.

From the first, competitions were open to all, and the first challenges could be accomplished. The cost of entry was relatively low: those who had robots used them, those who didn't played in the simulation league. The first tabletop games were played on a misshapen pitch—a common ping-pong table—so participants would not have to build special tables. Although robotic soccer seems to offer an endless series of research challenges, its evaluation criterion is familiar to any child: win the game! The competitions are enormously motivating and bring in thousands of spectators (for example, 150,000 at the 2004 Japan Open). Two hundred Junior League teams participated in the Lisbon competition, helping to ensure robotic soccer's future.

It isn't all fun and games: RoboCup teams are encouraged to submit technical papers to a symposium. The best paper receives the RoboCup Scientific Challenge Award.

Challenge 2: Handy Andy

As ABC News recently reported, people find ingenious ways to support themselves in college: "For the defenders of academic integrity, their nemesis comes in the form of a bright college student at an Eastern university with a 3.78 GPA. Andy—not his real name—writes term papers for his fellow students, at rates of up to $25 a page."

Here, then, is the Handy Andy challenge: *Produce a five-page report on any subject.* One can administer this test in vivo, for instance, as a service on the World Wide Web; or in a competition. One can imagine a contest in which artificial agents go against invited humans—students and professionals—in a variety of leagues or tracks. Some leagues would be appropriate for children. All the contestants would be required to produce three essays in the course of, say, three hours, and all would have access to the web. The essay subjects would be designed with help from education professionals, who

A defining feature of the Handy Andy challenge, one it shares with Turing's test, is its universal scope.

also would be responsible for scoring the essays.

As a challenge problem, Handy Andy has several good attributes, some of which it shares with robot soccer. Turing's test requires simultaneous achievement of many cognitive functions and doesn't offer partial credit to subsets of these functions. In contrast, robot soccer presents a *graduated series of challenges:* it gets harder each year but is never out of reach. The same is true of the Handy Andy challenge. In the first year, one might expect weak comprehension of the query, minimal understanding of web pages, and reports merely cobbled together from online sources. Later, one expects better comprehension of queries and web pages, perhaps a clarification dialog with the user, and some organization of the report. Looking further, one envisions strong comprehension and not merely assembly of reports but some original writing. The first level is within striking distance of current information retrieval and text summarization methods. Unlike the Turing test—an all-or-nothing challenge of heroic proportions—we begin with technology that is available today and proceed step-by-step toward the ultimate challenge.

Because a graduated series of challenges begins with today's technology, we do not require a preparatory period to build prerequisites, such as sufficient commonsense knowledge bases or unrestricted natural language understanding. This is a strong methodological point because those who wait for prerequisites usually cannot predict when they will materialize, and in AI things usually take longer than expected. The approach in Handy Andy and robot soccer is to *come as you are* and develop new technology over the years in response to increasingly stringent challenges.

The five-page requirement of the Handy Andy challenge is arbitrary—it could be three pages or ten—but the required length should be sufficient for the system to make telling mistakes. A test that satisfies the *ample rope requirement* provides systems enough rope to hang themselves. The Turing test has this attribute and so does robot soccer.

A defining feature of the Handy Andy challenge, one it shares with Turing's test, is its *universal scope.* You can ask about the poetry of Jane Austen, how to buy penny stocks, why the druids wore woad, or ideas for keeping kids busy on long car trips. Whatever you ask, you get five pages back.

The universality criterion entails something about evaluation: we would rather have a system produce crummy reports on any subject than excellent reports on a carefully selected,

narrow range of subjects. Said differently, the challenge is first and foremost to handle any subject and only secondarily to produce excellent reports. If we can handle any subject, then we can imagine how a system might improve the quality of its reports. On the other hand, half a century of AI engineering leaves me skeptical that we will achieve the universality criterion if we start by trying to produce excellent reports about a tiny selection of subjects. It's time to grasp the nettle and go for all subjects, even if we do it poorly.

The web already exists, already has near universal coverage, so we can achieve the universality criterion by making good use of the knowledge the web contains. Our challenge is not to build a universal knowledge base but to make better use of the one that already exists.

Challenge 3: Never-Ending Language Learning

Proposed by Murray Burke in 2002, this challenge takes up a theme of Lenat and Feigenbaum's (1987) paper "On the Thresholds of Knowledge." That paper suggested knowledge-based systems would eventually know enough to read online sources and, at that point, would "go critical" and quickly master the world's knowledge. There are no good estimates of when this might happen. Burke's proposal was to focus on the bootstrapping relationship between learning to read and reading to learn.

We always must worry that challenge problems reward clever engineering more than scientific research. Robot soccer has been criticized on these grounds. Among its many positive attributes, never-ending language learning presents us with some fascinating scientific hypotheses. One states that we have done enough work on the semantics of a core of English to bootstrap the acquisition of the whole language. Another hypothesis is that learning by reading provides sufficient information to extend an ontology of concepts and so drive the bootstrapping. Both hypotheses could be wrong; for example, some people think that the meanings of concepts must be grounded in interaction with the physical world and that no amount of reading can make up for a lack of grounding. In any case, it is worth knowing whether one can learn what one needs to understand text from text itself.

Challenge 4: The Virtual Third Grader

One answer to the question, "if not the Turing test, then what?" was suggested by David Gunning in 2004: If we cannot pass the Turing test today, then perhaps we should set up a "cognitive decathlon" or "qualifying trials" of capa-

bilities that, collectively, are required for Turing's test. Howard Gardner's inventory of multiple intelligences is one place to look for these capabilities. However, it isn't clear how to test whether machines have them. Another place to look is elementary school. Every third-grader is expected to master the skills in table 1. All of them can be tested, although some tests will involve subjective judgments. Here is what my daughter wrote for her "convincing letter" assignment:

> Dear Disney,
>
> It disturbs me greatly that in every movie you make with a dragon, the dragon gets killed by a knight. Please, if you could change that, it would be a great happiness to me. The Dragon is my school mascot. The dragon isn't really bad, he/she is just made bad by the villan [sic]. The dragon is not the one who should be killed. For example, Sleeping Beauty, the dragon is under the villaness's [sic] power, so it is not neccisariliy [sic] bad or evil. Please change that.
>
> Your sad and disturbed writer,
>
> *Allegra.*

Although grading these things is subjective, there are many diagnostic criteria for good letters: The author must assert a position (stop killing the dragons) and reasons for it (the dragon is my school mascot, and dragons aren't intrinsically bad). Extra points might be given for tact, for suggesting that the recipient of the letter isn't malicious, just confused (the dragon isn't the one who should be killed, you got it wrong, Disney!)

Criteria for Good Challenges

You, the reader, probably have several ideas for challenge problems. Here are some practical suggestions for refining these ideas and making them work on a large scale. The success of robot soccer suggests starting with easily understood long-term goals (such as beating the human world soccer team) and an organization whose job is to steer research and development in the direction of these goals. The challenge should be administered frequently, every few weeks or months, and the rules should be changed at roughly the same frequency to drive progress toward the long-term goals.

The challenge itself should test important cognitive functions. It should emphasize comprehension, semantics, and knowledge. It should require problem solving. It should not "drop the user at approximately the right location in information space and leave him to fend for himself," as Edward Feigenbaum once put it.

A good challenge has simple success criteria.

The challenge itself should test important cognitive functions. It should emphasize comprehension, semantics, and knowledge. It should require problem solving.

Understand and follow instructions

Communicate in natural language
(for example, dialog)

Learn and exercise procedures
(for example, long division, outlining a report)

Read for content
(for example, show that one gets the main points of a story)

Learn by being told
(for example, life was hard for the pioneers)

Common sense inference
(for example, few people wanted to be pioneers)
and learning from commonsense inference

Understand math story problems and solve them correctly

Master a lot of facts (math facts, history facts, and so on).
Mastery means using the facts to answer questions
and solve problems.

Prioritize
*(for example, choose one book over another,
decide which problems to do on a test)*

Explain something
(for example, why plants need light)

Make a convincing argument
(for example, why recess should be longer)

Make up and write a story about an assigned subject
(for example, Thanksgiving)

Table 1. Third-Grade Skills (thanks to Carole Beal).

However an attempt is scored, one should get specific, diagnostic feedback to help one understand exactly what worked and what didn't. Scoring should be transparent so one can see exactly why the attempt got the score it did. If possible, scoring should be objective, automatic, and easily repeated. For instance, the machine translation community experienced a jump in productivity once translations could be scored automatically, sometimes daily, instead of subjectively, slowly, and by hand.

The challenge should have a kind of monotonicity to it, allowing one to build on previous work in one's own laboratory and in others'. This "no throwaways" principle goes hand-in-hand with the idea of a graduated series of challenges, each slightly out of reach, each providing ample rope for systems to hang themselves, yet leading to the challenge's long-term goals. It follows from these principles that the challenge itself should be easily modified, by changing rules, initial conditions, requirements for success, and so on.

A successful challenge captures the hearts and minds of the research community. Popular games and competitions are good choices, provided that they require new science. The cost of entry should be low; students should be able to scrape together sufficient resources to participate, and the organizations that manage challenges should make grants of money and equipment as appropriate. All participants should share their technologies so that new participants can start with "last year's model" and have a chance of doing well.

In addition to these pragmatic and, I expect, uncontroversial suggestions, I would like to suggest three others which are not so obviously right.

First, Turing proposed his test to answer the question "Can machines think?" but this does not mean a challenge for AI must provide evidence for or against the proposition that computers have intentional states and behaviors. I do not think we have any chance of testing this proposition. There are no objective characterizations of human intentional states, and the states of machines can be described in many ways, from the states of registers up to what Newell called the knowledge level. It is at least technically challenging and perhaps impossible to establish correspondences between ill-specified human intentional states and machine states, so the proposition that machines "have" intentional states probably cannot be tested. Perhaps the most we can require of challenge problems is that they include tasks that humans describe in intentional terms.

Second, in any given challenge, we should accept poor performance but insist on universal coverage. I admit that it is hard to define universal coverage, but examples are easily found or imagined: Reading and comprehending any book suitable for five year olds; producing an expository essay on any subject; going up the high street to several stores for the week's shopping; playing Trivial Pursuit; creating a reading list for any undergraduate essay subject; learning classifiers for a thousand data

sets without manually retuning the learner's parameters for each; playing any two-person strategy game well with minimal training; beating the world champion soccer team. Each of these problems requires a wide range of capabilities, or has a great many nonredundant instances, or both. One could not claim success by solving only a part of one of these problems or only a handful of possible problem instances. What should we call a program that plays chess brilliantly? History! What should we call a program that plays any two-person strategy game, albeit poorly? A good start! A program that analyzes the plot of Romeo and Juliet? History! A program that summarizes the plot of any children's book, albeit poorly? A good start! Poor performance and universal scope are preferred to good performance and narrow scope.

My third and final point is related to the last one. Challenge problems should foster what I'll call a *developmental* research strategy instead of the more traditional and generally successful *divide and conquer* strategy. The word *developmental* reminds us that children do many things poorly, yet they are complete, competent agents who learn from each other, and adults, and books, and television, and playing, and physical maturation, and other ways, besides. In children we see gradually increasing competence across many domains. In AI we usually see deep competence in narrow domains, but there are exceptions: robotic soccer teams have played soccer every year since the competitions began. If the organizers had followed the traditional divide and conquer strategy, then the first few annual competitions would have tested bits and pieces—vision, navigation, communication, and control—and we probably would still be waiting to see a complete robotic team play an entire game. Despite the success of divide-and-conquer in many sciences, I don't think it is a good strategy for AI. Robotic soccer followed the other, developmental strategy, and required complete, integrated systems to solve the whole problem. Competent these systems were not, but competence came with time, as it does to children.

Conclusion

In answer to the question, "if not the Turing Test, then what," AI researchers haven't been sitting around waiting for something better; they have been very inventive. There are challenge problems in planning, e-commerce, knowledge discovery from databases, robotics, game playing, and numerous competitions in aspects of natural language. Some are more successful or engaging than others, and I have discussed some attributes of problems that might explain these differences. My goal has been to identify attributes of good challenge problems so that we can have more. Many of these efforts are not supported directly by government, they are the efforts of individuals and volunteers. Perhaps you can see an opportunity to organize something similar in your area of AI.

Acknowledgments

References

Dennett, D. 1998. Can Machines Think? In *Brainchildren, Essays on Designing Minds.* Cambridge, MA: MIT Press.

French, R. 2000. The Turing Test: The First Fifty Years. *Trends in Cognitive Sciences* 4(3): 115–121.

Gardner, H. 1983. *Frames of Mind: The Theory of Multiple Intelligences.* New York: Basic Books.

Hayes, P., and Ford, K. 1995. Turing Test Considered Harmful. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence,* p. 972. San Francisco: Morgan Kaufmann Publishers.

Lenat, D. B., and Feigenbaum, E. A. 1987. On the Thresholds of Knowledge. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence,* 1173–1182. San Francisco: Morgan Kaufmann Publishers.

Mackworth, A. 1993. On Seeing Robots. In *Computer Vision: Systems, Theory and Applications,* ed. A. Basu and X. Li, 1–13. Singapore: World Scientific Press, 1993. Reprinted in *Mind Readings,* ed. P. Thagard. Cambridge, MA: MIT Press, 1998.

Turing, A. 1950. Computing Machinery and Intelligence. *Mind* 59(236):433–460.

Whitby, B. R. 1996. The Turing Test: AI's Biggest Blind Alley? In *Machines and Thought: The Legacy of Alan Turing, Vol. 1.,* ed. P. Millican and A. Clark. Oxford: Oxford University Press. (www.informatics.sussex.ac.uk/users/blayw/tt.html.)

Paul Cohen is with the Intelligent Systems Division at USC's Information Sciences Institute and he is a Research Professor of Computer Science at USC. He received his Ph.D. from Stanford University in computer science and psychology in 1983 and his M.S. and B.A. in psychology from the University of California at Los Angeles and the University of California at San Diego, respectively. He served as a councilor of the American Association for Artificial Intelligence (AAAI) from 1991 to 1994 and was elected in 1993 as a fellow of AAAI.