

What Has the Loebner Contest Told Us About Conversant Systems?

Kenneth R. Stephens
Operant WebSites, Inc.

The [Standard Turing Test](#), proposed by [Alan Turing](#) in 1950 as a test of whether machines can actually think, is an exercise in making human judges think a computer program is actually a human. The annual competition for the [Loebner Prize](#), which has been held every year since 1991 (with the Cambridge Center for Behavioral Studies as the original sponsor), was intended as the “first instantiation of the Turing Test.” No competitor has yet come close to winning the \$100,000 Grand Prize, but each year a smaller award and the Bronze Prize are awarded to the most human-like contestant.

After eleven years of contests, it is appropriate to look back at the most successful systems of the past, and in doing so, try to answer the following questions:

1. What have been the most successful techniques?
2. To what extent have these mirrored the state-of-the-art in computational linguistics and natural language processing?
3. Are we any closer to a truly intelligent computer, and is the Loebner competition the right way to find out?

Contents:

- [Non-Loebner Conversant Systems](#)
- [Loebner Contest Winners](#)
- [Conclusions](#)
- [References](#)

Non-Loebner Conversant Systems

We must begin our review of conversant systems several years before the first

Loebner competition, as some of the winners were descendants of two early programs — ELIZA and PARRY — and the SHRDLU system is a benchmark for any intelligent conversational system.

ELIZA

ELIZA was a program written by Joseph Weizenbaum ([Weizenbaum, 1966](#)) from 1964 to 1966 while he was a researcher at M.I.T. ELIZA carried on a “conversation” in probably the most limited manner possible — by reacting as though it were a Rogerian therapist, it simplified its task considerably. Psychotherapists trained in the tradition of Carl Rogers tried to be “nondirective.” Rather than asking leading questions that might steer the dialog in the direction of the topics that interested Freudian and Jungian psychoanalysts (“Tell me more about your sexual feelings toward your mother”), Rogerians simply turned the clients' statements around to show some comprehension but to elicit more statements from the client which might eventually lead them to an understanding of their own conflicts and problems. (In practice, a videotape analysis by Truax (1966) showed that even Carl Rogers was directing the dialog through subtle verbal reinforcement of themes he found interesting.) At any rate, a Rogerian therapist did not have to initiate new themes in the conversation, and as the ELIZA program demonstrated, did not have to exhibit any knowledge or understanding in the traditional sense. It only had to maintain the illusion of carrying on its end of the conversation, which made it an ideal candidate for a system given the limited capabilities of computer programs in the 1960s.

The computational task of ELIZA was

- 1) parse the input from the user to find keywords, and to identify the most important keywords to which it should respond
- 2) use rules and templates to transform the input (turn “I” into “you,” wrap key

statements in such contexts as “in what way do you think ___”, “does it please you to believe ___”, “how long have you been ___”, etc.) into a response that echoed the input statement or asked for more discussion around the keyword

3) if the user input contained no identifiable keywords or content, generate a reasonable response.

There was no overall script to the conversation, other than a list of transformational and decomposition rules — ELIZA was purely reactive.

Weizenbaum was surprised and dismayed by the extent that humans accepted ELIZA without question. In his 1976 book, he pointed out some of the concerns that led him to drop this line of research: psychoanalysts believed that this was the basis of an acceptable, automated therapeutic regime in the future, users quickly became emotionally attached to the program, and some learned people felt that ELIZA had demonstrated a general solution to man-machine communication. To some extent, perhaps these reactions reflected the lack of general understanding by most people at that time of the capabilities and limitations of computers.

PARRY

A few years after ELIZA was written, a group of computer scientists at Stanford AI Lab led by Kenneth Colby produced a system called PARRY that simulated the dialog between a therapist and a paranoid schizophrenic. PARRY modeled a paranoid with distorted belief systems and inability to stay on topic for very long, and was so successful it was used to train psychologists, many of whom were unable to distinguish PARRY's dialog from that of a real paranoid patient. In that sense, it might be said that PARRY satisfied a limited form of the Turing Test even in the mid-1970's.

PARRY was more challenging to develop than ELIZA, because it could not simply reflect what the user had said, but had to have something to say as well. However, since the output was often gibberish or a non sequitur, it might fool psychologists into thinking they were interacting with a schizophrenic, but would not do well in a test in mimicking a normal conversation. Nevertheless, the PARRY

system influenced some of the later Loebner contestants. In fact, a version of PARRY competed in the early Loebner contests.

SHRDLU

In 1972, Terry Winograd published his doctoral research at MIT on the SHRDLU system. SHRDLU simulated the actions of a robot interacting with a “blocks world” of different colored and shaped blocks which could be placed on a table or put in a box. It maintained an internal model of the state of the different blocks (where they were, their position relative to other objects, what manipulations had been done to them lately, etc.). SHRDLU could carry on a conversation about the blocks world with its user, and was very sophisticated as far as sentence parsing of input commands and questions (its replies, however, were fairly limited and even boring.) In the sense that it answered correctly about the status of the blocks world, it showed what could be called limited understanding of its virtual environment. It is indeed impressive to examine some of the early logs, as it appears that the level of discourse supported by SHRDLU is at a much higher level than contemporary systems. (However, it is important to note that if the commands or questions had been phrased a little differently, the system might not have appeared so robust. It is also important to note that it was relatively ignorant of the very things it was supposed to know about, so that it had no knowledge that blocks were heavy, that they dropped if moved off the table, that larger blocks weighed more than smaller, etc.)

SHRDLU can be considered a knowledge-based system. It employs “procedural representations” of words and sentences, so that functions written in LISP and MICRO-PLANNER embodied the meaning of a command or question, and the reasoning capabilities of SHRDLU existed within these programs.

A parsing module, called PROGRAMMAR, was written to implement an approach to “systemic grammar” which was a different method of parsing than methods like backtracking (in which grammar rules are applied with identification of choice points, and if parsing is blocked, the process backtracks to the choice point and tries a different rule) or parallel processing (in which many different possible

structures are built concurrently, and then the best is chosen). The details of systemic grammar, as well as concepts such as context-free grammars, augmented transition networks, semantic networks, knowledge representation schemes such as frames, rules, etc., and parsing strategies are beyond the scope of the present paper, but the interested reader can find background on these techniques in artificial intelligence texts (e.g. the chapter on “Understanding Natural Language” in Barr and Feigenbaum (1981). SHRDLU employed much more powerful techniques than did ELIZA.

Loebner Contest Winners

Hugh [Loebner has said](#) “I believe that this contest will advance AI and serve as a tool to measure the state of the art.” By contrast, in a review which was critical of the Loebner Prize, [Shieber \(1994\) said](#) “...the Loebner prize competition neither satisfies its own avowed goals [of “pushing the envelope”], nor the original goals of Alan Turing.” Whatever was true in 1994, let us review the history of the entire competition and its winners to see whether the competition has become more challenging, and to what extent it reflects the state of the art.

PC Therapist

The winner of the first Loebner competition in 1991 was a program called PC Therapist III, written by [Joseph Weintraub](#). It won in the category of “whimsical conversation,” which was perfect for the kind of *non sequitur* it would deliver. Several of the later winners of the competition have dismissed Weintraub's system as nothing more but an ELIZA clone, but in truth, Weintraub was so disappointed in the version of ELIZA which he had purchased, that he returned it and set about to write his own. Weintraub won the competition in 1991, 1992, 1993, and again in 1995. Two of his other systems were called PC Professor (which was prepared to talk about

the differences between men and women) and PC Politician (which discussed liberals and conservatives.)

PC Therapist was touted to employ “AI sentence parsing and knowledge base technology.” If it could not construct a reply based on keyword parsing, it would pull a relevant quote or phrase from its knowledge base, called KBASEK.

TIPS

In 1994, the winning program was written by [Thomas Whalen](#) of the Communications Research Centre, in Ottawa, Canada. His TIPS system was designed to provide information rather than simulate a conversation; and, to that end, referenced a database of pre-written answers to specific questions. Whalen regards himself as “Computational Behaviorist,” and his TIPS system relied more on an analysis of actual human behavior, than on a linguistic-based grammar.

[For the 1995 competition](#), he tried to equip his entry with the personality of a specific person to cope with new rules that opened up the conversation to multiple topics (up through 1994, an entry could choose to carry on a conversation in a limited domain). He hoped that by engaging the judges in a dialog around a story of the trials and tribulations of TIPS' character (“Joe the Janitor”), who was about to lose his job; thereby constraining the conversation to a pre-scripted domain while being human-like. He also programmed Joe to deal with the kinds of questions Dale Carnegie said usually came up first in polite conversation. He was wrong. The judges did not engage in polite conversation and were not interested in playing along with the Joe the Janitor scenario; instead they immediately tried to confound the system and identify it as an imposter. He also found that the judges were less impressed by his system's four variations on the standard answer of “I don't know” than they were of a system that provided witty *non sequiturs* instead. These observations proved to be helpful to later contestants.

Whalen recently made the following observation:

“Looking at grammars immediately forces one to wrestle with some of the most intractable problems in linguistics... I believe that approaching the problem from the direction of verbal

behaviour rather than grammatical analysis provides a less steep learning curve for the program developer. By starting with small samples of behaviour and generalizing to more difficult and larger samples, we can develop useful systems earlier and avoid the seemingly insurmountable obstacles that engage and confound linguists from the outset. Most linguists, of course, do not believe that a behaviour-oriented approach will even encounter the most fundamental problems of linguistics much less solve them. The great thing about the Loebner competition is that it allows programs from any philosophical approach to compete head-to-head without prejudice or bias. We are seeing the competition bear fruit — the competing programs (even the computational behaviourists' programs) are becoming sufficiently sophisticated to begin to address some of the fundamental problems of language understanding. But it is still too early to see which approach will win in the long run.” (Whalen, 2002, personal communication.)

MegaHAL and HeX

In 1996, the primary motivation for Jason Hutchens, of the University of Western Australia, was to [make a statement about the “futility of the Loebner contest.”](#) He intentionally limited his development time to one month and incorporated no technologies from Artificial Intelligence. His opinions, which he has posted on the web, are that the Loebner competition does not attract serious attention from the A.I. community, and that it does nothing to “push the envelope” of modern technology. If he could win with a one-month “hack” it would show that the Loebner competition is irrelevant to anyone doing serious A.I. work. He was quoted as saying that his systems had nothing to do with artificial intelligence, and in fact, were about “as smart as a Mr. Coffee.”

Hutchens entered two systems in 1996: MegaHAL and HeX. At that time, MegaHAL was nothing more than a gibberish-generator, although in 1998 it had been improved enough to be Hutchens' primary entry. It makes use of third order Markov chains that relate the probabilities of one word pair given another preceding word pair in a state transition model, and then it makes use of as many keywords as it can from the user's input to make it look like it is conversing. The probabilities were derived from the transcripts of previous Loebner judges. MegaHAL at this time was not the serious entry — that was HeX, which made use of MegaHAL as just one of its modules.

Like Whalen, he equipped HeX to talk about the “Dale Carnegie” polite conversation starters, but he anticipated that the judges would also probably introduce weird topics. He also found that sometimes the judges would type in several

sentences in a row, so HeX was able to react to more than just the last one. He found that the judges often asked questions that started with “wh_” (what, when, where, why, who, etc.) and they almost always ended such questions with a question mark. So it was easy to parse questions that could not be answered and reflect them back as a statement. He also found that the dead giveaway for some entrants had been that they sometimes repeated, verbatim, the same response, but that humans did not.

The method of constructing an answer was to iterate roughly in this order:

- Parse sentences one-by-one, convert to words. Look for keywords in a database of hardwired replies (and use one only if hadn't been used before).
- If a stored reply could not be located, evaluate for a trick question, and if detected, give a witty reply.
- Call MegaHAL and generate psychobabble.
- Reformulate the user's input according to one of several hundred templates and spit it back.
- Give a humorous response to silence.
- Accuse the user of being ungrammatical etc.
- As a last resort, generate more psychobabble with MegaHAL.

In 1997, he entered a “considerably more powerful program, SEPO” but lost to David Levy's team, and in 1998, he entered the [reworked MegaHAL](#) but again failed to win. He attributes this to an increase in the sophistication of the competitors.

CONVERSE (Catherine)

Part of the reason that the competition began getting tougher was the resurgence of an empirical approach within the field of computational linguistics in the mid-1980s and the related development of a family of probabilistic techniques at places like the IBM Thomas J. Watson Research Center. According to Manning and Schütze (1999), empiricism had been the driving force in language research in the early days until the ascendancy of “rationalism” in the early 1960s. Chomsky and others believed that the logic and rules of grammar determine an utterance, and that these rules are innate. However, this point of view has gradually fallen out of favor, and a more stochastic paradigm is now more credible. An empirical point of view sees the way that words are used as changeable according to the practices of the verbal

community, and it more easily accounts for dialects and gradual shifts in language than does a rigid set of rules. “Meaning” exists not within the word, but within the way that people speak and understand the word.

Within the context of this “sea change” in the field of natural language processing, a team from Yorkshire, England, representing Sheffield University and Intelligent Research, Inc. and led by David Levy, put together a conversant personality they called Catherine, based on the [CONVERSE system](#), for the 1997 competition. Like the old PARRY system, Catherine took the initiative in the conversation rather than being a passive participant. The idea was that Catherine would control the conversation, allowing fewer opportunities for the judges to ask unconstrained questions. At the same time that Catherine's scripts were trying to control the direction of the conversation, a module to respond to questions and comments accessed a database of information relevant to the conversation. These databases included a thesaurus and a dictionary of proper names. They also included a Person database which had all the relevant biographical information about Catherine's character and other characters such as her fictitious family and friends. A weighting system could control which of these modes (the top-down script, or the bottom-up question-answering module) would predominate depending on how the conversation was proceeding.

A sophisticated text parser had been trained using the statistics from analyses of a corpus of British dialog. One can say that in general, Catherine took better advantage of current speech and language tools than her predecessors. Catherine carried on a conversation with the judges that was heavily laced with current events about Bill Clinton, Whitewater, and the “coming out” of two lesbians at the White House the night before the contest. Probably the currency of her conversation topics played a part in making her human-like.

Since winning the 1997 Loebner Prize, Levy has focused on investigating conversant agents for small platforms and embedded systems.

FRED, Barry DeFacto, Albert, and Albert2

[Robby Garner](#), like Thomas Whalen, considers himself a computational behaviorist, continuing the trend toward an analysis of the dialogues with previous judges. His systems won the competition two years in a row (1998 and 1999) and he has remained active with the Loebner competition, giving technical support to current contestants.

[Garner has beliefs](#) that resonate strongly with many people who consider themselves behaviorists:

Human intelligence is such a vague term, and we don't fully understand ourselves. Artificial Intelligence has become even more vague, and I like to say that I don't believe in the word intelligence at all because human behavior is more than just the meanings of our words and how many of them you can remember.

...

I define intelligence as the capacity to acquire and apply knowledge. Knowledge is familiarity, awareness, or understanding gained through experience or study.

...

Passing the Turing test does not require intelligence.

Garner's systems that won the Loebner competition matched phrases they have seen in the past, and when new phrases were encountered, they were flagged for later refinement. Conversations were modeled as Stimulus-Response objects, and where possible, appropriate and humorous answers were associated with questions. However, there were good backup strategies in case there was not an existing association. For instance, after parsing the input, an algorithm looked at frequencies of the words used and found the three most significant words. It then performed “database mining” (from a huge database that included Probert's Electronic Encyclopedia, the Jargon File, The Devil's Dictionary and other references) and put the data returned into one of many response templates.

ALICE

Richard S. Wallace won the next two Loebner competitions (2000 and 2001) with [ALICE](#) (an acronym for Artificial Linguistic Internet Computer Entity.) ALICE

is based on the [Artificial Intelligence Markup Language \(AIML\)](#). Wallace, like several other contestants before him, is very open in characterizing his systems as based on a “strategy of deception and pretense... [that] can be traced through the history of artificial intelligence.”

A relatively small number of important constructs are important in AIML:

- A <category> contains a <pattern> (which is the stimulus) and a <template> (which is the response)
- The stimulus-response categories are stored in a tree structure, managed by an object called Graphmaster
- The <srai> recursion tag is used for several purposes: symbolic reduction; “divide and conquer” (splitting an input into subparts, and combining responses to each of these); finding synonyms; correcting spelling or grammar; and/or conditional branching.
- <that> is a contextual variable that encloses a pattern based on ALICE's previous responses
- <topic> collects a group of <categories>
- <person> provides a pronoun-swapping function that transforms part of the input for inclusion in the output.
- An iterative process of supervised learning, called “targeting”, flags cases in which there are no matching templates to a <pattern> so that the “botmaster” can manually intervene and write an appropriate response.

Wallace sees [ALICE as in the same family as ELIZA](#) but more sophisticated by many degrees. [He says](#) that you do not need artificial intelligence to pass the Turing Test, nor do you need “complex theories of learning, neural nets and cognitive models.”

EllaZ

For the 2002 competition, Kevin Copple entered a simplified version of the [EllaZ program](#), called simply “Ella.” Copple made good use of the best techniques of previous winners, including those of ALICE / AIML. He says “my approach is to use tools that are available, be clever where I can, do a lot of work, find ways to re-use work of others, and prepare for more sophisticated approaches.”

When asked what makes Ella different than other ELIZA-like approaches, he listed the following techniques:

- Extensive word math such as “what is ten point three over a dozen added to a grand” including trigonometry, logarithms, and exponents.
- Check for gibberish such as pounding on the keyboard or speaking in a foreign language. Check for at least two actual words from a scrabble dictionary when the sentence is long enough to test.
- Use WordNet database as a backup to answer test questions such as “what is a horse” or “where is Spain” by just getting the most common definition and putting it in a natural speaking format such as “I happen to know that ...”
- Do “monologues” where Ella ignores the visitor for a few lines and just says “what she wants.” Allow function type questions to interrupt, such as “what is 29C in F?” but then Ella goes back to the monologue.
- A method I call “Convuns” or conversational units. These are a large collection of poems, jokes, trivia, maxims, fables, limericks, and so on.
- Take the user's words and reuse without the tense/pronoun shifts that can also be effective. Such as “‘What next' you say? Wow, THAT's original.”
- Prepare for common questions like capital of, color of, using humor when a specific answer is not available.
- Develop and reuse functions such as extracting yes/no or extracting a number.
- Avoid having Ella say wrong things. She will not answer “Should I ...” with “yes, definitely” as some other chatterbots will.

Conclusions

We have reviewed the techniques, at a very high level, of the winners of the annual

Loebner competition, as well as some of their predecessors such as ELIZA, PARRY and SHRDLU. We have seen that the technical sophistication and capabilities of the winners

have increased, particularly in the last 5 years or so. One can still see the influences of ELIZA and PARRY in current day winners, although by degree, the Loebner winners are significantly more capable. What we have not seen is very much influence by the artificial intelligence techniques embodied in SHRDLU. This has led certain of the contestants to conclude that the Loebner competition is of no relevance to AI. But we have seen the influence of current trends in natural language processing away from “rationalist” methods back to a more “empirical” approach.

One might be tempted to state at this juncture that the Standard Turing Test does not demonstrate what Alan Turing thought it would. Block (1990) has tried to make exactly this argument. The Loebner contest, as “the instantiation of the Turing Test,” rewards deception and pretense. It does not require true “intelligence” of the sort referred to by Garner (i.e. the acquisition and application of knowledge) so some of Hutchens' criticisms of the Loebner contest seem justified. On the other hand, contrary to his assertions, the bar has been raised considerably over the last few years. An ELIZA-like approach is no longer sufficient. Techniques are being used now to statistically model a conversational corpus and to generate responses by Markov models and other sophisticated algorithms. With the number of serious entries increasing every year (there were six in 1991 and there were 31 in 2002) competition is increasing so that level of capabilities shown by recent winners will no longer be adequate.

We have seen that it is important to constrain the judges from exploring unanticipated tangents. This has been done with scripts and character development that engage the judge in talking about a particular topic. Another essential ingredient, however, is to be able to gracefully handle the judges' statements or questions when they go “off topic.” To this end, online resources such as encyclopedias, thesauruses, databases of names and characteristics of the character and those in its fictitious world, all help construct a response that keeps the conversation going. Stochastic methods of constructing plausible sentences have shown their worth. It is also important not to repeat the same answer verbatim, as this is a certain indicator that the persona is software, so systems with hundreds of acceptable ways of saying the same

thing have a clear advantage. The witty *non sequitur* seems more human-like to the judges than a simple “I don't know.” In general, a careful analysis of the dialogs from previous contests has been helpful to the winners.

Richard Wallace said that you do not need artificial intelligence to win the Loebner prize. That may still be true today. But [new computational models of verbal behavior](#), based on adaptive models such as neural networks and incorporating Skinner's functional analysis, have recently been developed (reference Bill's paper on the site) and will raise the bar even further. Alan Turing himself made this comment in his 1950 paper: “It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and teach it to understand and speak English. This process could follow the normal teaching of a child.” Perhaps a system whose world view does not consist of witty *non sequiturs*, “canned” replies, and references to a database of facts, but which has been learned through interactions with humans and is situated in the real world, is exactly what will be required to win the \$100,000 Loebner Prize. It will be exciting to see the kind of systems that emerge in the future to challenge for the Loebner Prize. David Levy (2002, personal communication) feels the same way — he says “In my view the next few years will see a plethora of sophisticated, commercially available conversant agents in use in a variety of domains, ranging from entertainment to medicine, law and the other professions.”

The trends in the field of computational linguistics parallel a similar shift in the field of artificial intelligence from purely symbolic, rule-based models to distributed, probabilistic models such as neural networks. The paradigm shifts in both fields brings them into compatibility with B.F. Skinner's (1957) analysis of verbal behavior, and elsewhere on this web site, you will find papers ([MacCorquodale 1970](#), [Catania 1972](#), [Palmer 1987](#) and [Palmer 2000](#)) pointing out shortcomings of Chomsky's curt rejection of Skinner's analysis and his nativist philosophy. The fact that other fields are also seeing the flaws in Chomsky's system, including the circularity of unprovable concepts such as the “Language Acquisition Device,” provides some vindication for

Skinner's 1957 analysis, and reason for those fields to reexamine the functional analysis contained in *Verbal Behavior* and work that subsequently has extended it.

References

- Barr, A. and Feigenbaum, E. A. (1981). *The Handbook of Artificial Intelligence*. Los Altos, CA: William Kaufmann, Inc.
- Block, N. (1990). The computer model of the mind. In D. N. Osherson and E. E. Smith (Eds.) *An Introduction to Cognitive Science III: Thinking*. Cambridge MA: MIT Press, pp. 147-289.
- Colby, K. M. (1964). Experimental Treatment of Neurotic Computer Programs. *Archives of General Psychiatry*, 10: 220-227.
- Colby, K. M. (1975). *Artificial Paranoia: A Computer Simulation of Paranoid Processes*. New York: Pergamon Press.
- Manning, C. D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: The MIT Press.
- Skinner, B. F. (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.
- Truax, C.B. (1966). Reinforcement and non-reinforcement in Rogerian psychotherapy. *Journal of Abnormal and Social Psychology*, 71, 1-9.
- Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, 59, 433-460.
- Weizenbaum, J. (1966). ELIZA - A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9, 36-45.
- Weizenbaum, J. (1976). *Computer Power and Human Reason*. W. H. Freeman and Company.
- Winograd, T. (1972). *Understanding Natural Language*. New York: Academic Press.

Links to Selected On-line Resources

Background: The Turing Test and A.I.

[The Turing Test Page](#)

[Computing Machinery and Intelligence A.M. Turing](#)

[IMHOFAQ: The Outsider's Guide to Artificial Intelligence](#)

[Machine Language, by Tracy Quan](#)

Pre-Loebner Systems

[Joseph Weizenbaum's ELIZA: Communications of the ACM, January 1966](#)

[ELIZA \[Weizenbaum 1966\]](#)

[PARRY Encounters The Doctor](#)

[PARRY](#)

The Loebner Competition

[Loebner Prize Home Page](#)

[Machine Intelligence, Part I: The Turing Test and Loebner Prize, by Ashley Dunn](#)

[1995 Loebner Prize Information](#)

[1996 Loebner Prize Information](#)

[1997 Loebner Prize Information](#)

[Loebner Atlanta 2002](#)

[Lessons from a Restricted Turing Test, by Stuart M. Shieber](#)

[In Response, by Hugh Loebner](#)

Loebner Winners

[Joseph Weintraub Information](#)

[Dr. Thomas Whalen's Biography](#)

[Thom's Participation in the Loebner Competition 1995, or How I Lost the Contest and Re-Evaluated Humanity, by Thom Whalen](#)

[Jason Hutchens: how megahal works](#)

[How to Pass the Turning Test by Cheating, by Jason L. Hutchens](#)

[CONVERSE](#)

[Robby Glen Garner](#)

[\[Robby Garner's\] bio](#)

[How I Failed The Turing Test Without Even Being There, by Robby Garner](#)

[A.L.I.C.E. AI Foundation](#)

[AIML Overview, by Richard Wallace](#)

[The Anatomy of Alice, by Richard Wallace \(slideshow\)](#)

[From ELIZA to A.L.I.C.E., by Richard Wallace](#)

[A Conversation with Dr. Richard Wallace](#)

[EllaZ](#)