Social Interaction with Computers: an Interpretation of Weizenbaum's ELIZA and Her Heritage.

Hans Pruijt

Erasmus University Rotterdam

Submitted for publication in Social Science Computer Review (Pruijt, H., 2006, "Social Interaction with Computers: An Interpretation of Weizenbaum's ELIZA and Her Heritage", Social Science Computer Review, Volume 24, Number 4, 516-523)

Abstract
Social interaction with computers is a key part of the frontier of automation. Weizenbaum's ELIZA continues to be relevant as the focus of the debate shifts from artificial intelligence to the future of narrative and emotional design. A version of ELIZA was implemented in Excel/Visual Basic for Applications that combines adherence to the original specification with easy experimentation.

Keyword list: human computer interaction, automation, ELIZA, artificial intelligence, narrative

It is a truism that automation is perpetually on the rise. But where are the limits? This has been the subject of debate for decades. One of the most lasting and highly cited contributions to the debate is ELIZA, the pioneering conversation or chatterbot program originally developed by Weizenbaum (1966).
It played a role in the debate on Artificial Intelligence in the 1970s, especially as an example to show that Artificial Intelligence can hardly be more than simulated rather than emulated intelligence (Raggett and Bains 1992: 204). However, interest in ELIZA survived the AI hype. It surfaced in recent discussions on identity and on-line interaction (Turkle 1995), the future of narrative (Murray 1997) and emotional design (Norman 2004). ELIZA did not become a classic because of her technical sophistication. Kurzweil (1990: 36), for example, described her as a "simple minded program unrepresentative of the state of the art". ELIZA became timeless because she epitomizes a major section of the frontier of automation: social interaction with computers. Turing's (1950) acid test to determine whether machines can think involves social interaction with a computer. The Turing test is a game in which a test person (C) is engaged in two chats, one with a man (A) and one with a woman (B). C is asked to determine which chat partner is A and which chat partner is B. B's role is to help C make the correct identification; the role of A is to confuse C. Turing proposed to accept as proven that computers can think, when a computer could take over the role of A and be just as capable to trick C into making a wrong decision. ELIZA is a limited version of the Turing test (Weizenbaum 1966: 42).
The present paper explores how ELIZA informs the debate on social interaction with computers as the frontier of automation. The paper is complemented by a new ELIZA implementation, intended to open up ELIZA for scrutiny and in depth-experimentation. The ELIZA implementation is available at:
http://www.eur.nl/fsw/english/staff/homepages/pruijt/software

The original aim for producing the paper and the program was to add to the still very limited store of case studies that are suitable for use in a social science computing course. It may be seen as a sequel to an earlier case study on game theory (Pruijt 2000). The environment chosen for this implementation is Visual Basic for Applications (VBA) in MS Excel, mainly because of its widespread use and its orientation towards end users. Initially, I had hoped to find an existing implementation that I could convert into VBA. However, publicly available implementations written in suitable languages such as BASIC (North 1980: 253) did not, by any stretch of the imagination, come close to the original specification. Therefore, a completely new implementation was necessary. The goal was to write a program that had the same capabilities for analyzing the user's input and generating output that Weizenbaum (1966) had specified. The reason was avoid distortions caused by a simplistic understanding of ELIZA's workings. An added advantage of returning to the original specification was that it made it possible to refer to Weizenbaum's article for an additional explanation of key points in the program's design. (A very useful abridged version of the 1966 article is available on the web.) The paper includes some results from the experimental application of the case study in teaching.

How ELIZA works

Much of the literature conveys only a rudimentary notion of what ELIZA does. Raggett and Bains (1992: 231), for example, describe ELIZA as an "early NLP (natural language processing, H.P.) program, which mimicked a non-directive psychoanalyst by picking out key words in the typed input and printing out modifications of a stock of standard phrases."
While this is correct, ELIZA does more. Through pattern matching, it attempts to detect words of phrases in the input that define what the user is talking about; it makes use of this information to construct appropriate responses. I propose to take this as minimum criterion for stating that a program is designed to be a partner in social interaction. It sets ELIZA clearly apart from, for example, the natural-language help system present in MS Office (Miastkowski 1995). In terms of architecture, as in many AI applications in which processing logic and rule base are independent, there is a clean separation between the ELIZA program and the script. Weizenbaum's famous simulation of a psychotherapist was contained in a specially prepared script called "DOCTOR". For this case study, I have chosen a different target to simulate: a social research advisor.

A sample conversation (ELIZA's replies are in capitals):

HI. TELL ME ABOUT YOUR RESEARCH IDEAS
I want to study street violence
COULD YOU DESCRIBE THE PROBLEM OF STREET VIOLENCE IN THE SHAPE OF A CONCEPTUAL MODEL?
one independent variable is aggression
HOW DO YOU THINK TO MEASURE AGGRESSION ?
maybe there is a standard test
IT IS ALWAYS IMPORTANT TO CHOOSE THE RIGHT TEST
or a questionaire
COULD YOU TELL ME SOMETHING MORE ABOUT STREET VIOLENCE
it happens mainly after dark
LET US RETURN TO AGRESSION

often there is alcohol involved
GOOD
and it is a male affair
COULD YOU TELL ME SOMETHING MORE ABOUT THE SUBJECT
gangs are a big part of the problem
LET US TRY TO FOCUS OUR CONVERSATION A LITTLE BIT MORE

The basic operation that ELIZA performs is accepting an input sentence from the user and transforming it into an output sentence. This transformation involves four steps.

Step one: the program simply discards anything that was entered after the first (semi)colon.

Step two: ELIZA makes substitutions according to a search-and-replace table. A logical substitution is "YOU" for "I" and vice versa. The search-and-replace table is located on a worksheet in the ELIZA.XLS workbook. This makes it easy to change or add search-and-replace combinations.

Step three: the input sentence is broken down into its component parts. This process is governed by "decomposition rules". For example, one of these rules may entail the following: the first section will be the first word of the input sentence, the second section will be the word "THINK", and the third section will contain all the remaining words of the input sentence. Obviously, ELIZA can only apply this rule when the input sentence has "THINK" as its second word. ELIZA determines whether the rule applies by first looking if the word "THINK" exists in the input sentence, and if this is indeed the case, then by checking if the part of the input sentence that comes before THINK contains exactly one word.[1]

Step four: ELIZA assembles an output sentence. As building blocks, the program can use sections of the input sentence and also new material. This process is governed by "reassembly rules". For example, a reassembly rule might specify that the output will consist of the words WHY SHOULD IT BE THAT, followed by the third section of the original input sentence, and will be terminated by a question mark.

It is also possible to specify that one of the sections of the input sentence must be stored in one of four memory locations, for later use. This feature enhances the realism of the simulation.

To maintain compatibility with Weizenbaum's article, I have borrowed his notation for the decomposition rules. A digit (1-9) signifies a section with exactly so many words. The zero is used as a wildcard, which means a section with an arbitrary number of words. Thus, the decomposition rule in the example that I have used above will be written as `1THINK0.` This means: one (and not more than one) word, followed by the word "THINK", followed by an arbitrary number of words.
The notation for the reassembly rules also follows Weizenbaum. In this notation, digits refer to the order of the sections into which the input section was decomposed. Thus, the reassembly rule in the example that I have used above will be written as `WHY SHOULD IT BE THAT 3?.` This means "WHY SHOULD IT BE THAT", plus the third section of the decomposed input, plus a question mark.

Including the word `MEMORY` in a reassembly rule causes ELIZA to include the content of the top memory location in the output. An example of such a reassembly rule is:

`LET US GO BACK TO MEMORY`

Together, the search-and-replace table and the rules constitute the script. It should be an easy task to enter the rules in the spreadsheet. The lay-out is as follows:

Keyword	*Decomposition rule*	*Memorize*	*Reassembly rule*
THINK	1THINK0	3	WHY SHOULD IT BE THAT 3?

(In this example, the “3” under the heading “memory” means that the third section of the input should be committed to memory.)

Also in the ELIZA.XLS workbook, there is one worksheet for showing the recorded conversation, while a fourth worksheet can display a trace of all the operations performed by the program. The user can turn this option on by typing the word “tracing”.

A research advice script

The topic of the script may be anything, but a relatively realistic conversation should be possible when the context is chosen in such a way that it is logical for one party – the user – to supply all the substantive information to keep the conversation going. Nondirective therapy is one example of a suitable context, but for my experimental application in teaching I have chosen to set a different context: one in which a student - the user - who is planning to do some research, asks a lecturer - simulated by ELIZA plus script - for advice about how to proceed. The lecturer does not know anything about the subject of the research project, but tries to stimulate the student to clarify her or his thoughts. The lecturer has the option of using the context-free language of research methodology. This setting allows experimenting with the possibility of embodying some useful knowledge in the script, a deviation from the original that was purposely devoid of any meaning (Batacharia, Levy et al. 1999: 205). Also, the connection to previous training in research methodology adds another layer to the student’s experience.

In the experimental assignment, I presented the Research Advice Chatterbot as a joint project, analogous to an open source project. The starting point was the ELIZA spreadsheet equipped with a very rudimentary script. Students, working in small teams, were asked to contribute to its improvement. Acceptable contributions could have the shape of new combinations of keywords, decomposition rules and reassembly rules. Students had to specify where in the rule base the new rules should be inserted, because highly specific rules should be relative near the top; highly specific rules will hardly ever “fire” when placed below unspecific decomposition rules like `0IS0`.

Letting students write ELIZA scripts may be a novel application. Turkle (1995) described how ELIZA was used at MIT. In her account, students appear as passive users who did not write or modify scripts. The original implementation required would-be scriptwriters not only to understand the elementary notation for decomposition and reassembly rules, as is the case in the spreadsheet implementation, but also to master a more complex syntax. The method of entering the script on worksheets does away with this complexity.

Apart from script writing, other possibilities for student contributions to the Research Advisor Chatterbot were describing and analyzing bugs in the ELIZA program, and suggesting and implementing improvements. Integrating the contributions from the various student teams was just a matter of copy-pasting rules from students' spreadsheets into a master version. Actual contributions varied from adequate rules to cunningly designed rules, and in the case of one team, new VBA code aimed at improving the user-friendliness of the ELIZA program.

ELIZA and the limits of automation

One way to approach the frontier of automation is to try to identify what surely must be impossible to automate. ELIZA has proven to be a suitable starting point for such reasoning. The question is which functionality ELIZA critically lacks and whether such functionality can conceivably be expressed in the form of specifications for a computer program. Dreyfus (1992 :206) asserted that common sense knowledge about the world is essential to intelligent behavior. He pointed out that computers can only process such knowledge if we can reduce it to discrete, determinate elements and rules that define relationships between these elements. Dreyfus reasoned that humans do not think by manipulating elements and relationships, especially not when they are thinking about something that they have considerable experience with, and that decades of research in Artificial Intelligence had not yielded even the faintest notion of what common sense knowledge primitives should look like.

Reflecting on conversation programs, including ELIZA, Searle (1982) warned that we should not think that such programs can ever be made to understand anything. He illustrated his point by providing the example of a "Chinese Room", in which a non-speaker of Chinese is asked to process Chinese characters while following a set of clear instructions. [2] When Chinese observers will find the output indistinguishable from the output that could be produced by someone who does understand Chinese, this is equivalent to the Turing test being passed. However, few would claim that this implied that the person in the room understands Chinese.

The key ingredient that Searle maintained was missing in computers is intentionality. He saw intentionality as a unique feature of thinking as a biological process. He noted that it is peculiar that, although it is obvious that a computer simulation of photosynthesis cannot produce sugar, many are willing to believe that a simulation of the mind can produce intentionality.[3]

The alternative for trying to identify what is impossible to automate is to move towards the frontier of automation and see what is possible. Although the Turing test seems as elusive a goal as ever, chatterbots continue to be developed. Their main value seems to be entertainment. Recent experience from the Loebner Contest, a yearly competition for chatterbots, suggests that contenders that are in some way based on ELIZA can still impress the judges (Stephens 2002). In contrast with Searle's view, chatterbot developers Batacharia, Levy et al. (1999: 205) asserted that a program *can* embody intentionality. Indeed, compared to the DOCTOR script, the Research Advice Chatterbot embodies an attempt to incorporate some intentionality.

Social interaction with computers does not only depend on the characteristics of computer programs. The human capability for adaptation plays a role too. According to several authors, users' reactions to ELIZA revealed a "tendency to treat responsive computer programs as more intelligent than they really are" (Turkle 1995: 101).[4] Weizenbaum (1984) reported that users believed that the program could understand the information that they entered into the system. He witnessed his secretary, for example, confiding sensitive personal problems to the program. Hofstadter (1980: 600) called this phenomenon the "ELIZA Effect". Explanations advanced include the users' technical naivety[5], an increasingly mechanical world-view (Weizenbaum 1984), a general capability that people have to momentarily suspend disbelief (Turkle 1997: 103) and a penchant for anthromorphization (Norman 2004: 136).[6]

A Sisyphean struggle with the Turing Test is only one way to conceptualize social interaction with computers. We can also see it is as a development of interactive and multiform narrative. Murray (1997) introduced ELIZA as the world's first computer-based character, who demonstrated, for the first time in history, the "representational and narrative power" of the computer. In her analysis, the computer enhances a long-term trend for narrative to become more multiform, immersive and participatory. After the appearance of ELIZA, the trend continued with the rise of text based games that were later eclipsed by graphical games.

One application for which text based social interaction with computers seems suitable is humor. ELIZA shows potential for making fun of cliché-filled, meaningless discourse, precisely because of the rigidity of the computer (Murray 1997). Structuring interaction as a visit, to for example a therapist or advisor, is helpful. Murray (1997 :106) pointed out that "the visit metaphor is particularly appropriate for establishing a border between the virtual world and ordinary life because a visit involves explicit limits on both time and space."

Because of the easy way in which users can edit the script, we may see ELIZA.XLS as an "authoring system for interactive fiction writing" (Murray 1997: 297). We can frame the exercise of developing a Research Advice script as an attempt to develop an edutainment vehicle that could help thesis-writing students loosen up their imagination. The risk may be that it foreshadows thesis generators that constitute a new kind of hard-to-detect computer-assisted plagiarism.

A final twist in the ELIZA-inspired debate about what computer can do concerns emotions. Norman (2004) used ELIZA as an example to show that machines can induce emotions in people. He quotes the example of someone who was inadvertently chatting with ELIZA, while believing to be chatting with a real person, and quickly grew irritated as ELIZA was pointlessly keeping the conversation going. A question that remains is whether an ELIZA-like program can cause or reinforce positive emotions.

Hans Pruijt teaches at Erasmus University Rotterdam
Pruijt@fsw.eur.nl

References

Batacharia, B., D. Levy, et al. (1999). "CONVERSE: a Conversational Companion." in Y. Wilks (ed.) Machine Conversations. Boston, Kluwer: 205-214.

Dreyfus, H. L. (1992). What computers still can't do : a critique of artificial reason. Cambridge, MA, MIT Press.

Hofstadter, D. R. (1980). Gödel, Escher, Bach: An Eternal Golden Braid. New York, Basic Books.

Hofstadter, D. R. and D. C. Dennett (1982). "Reflections" in: D. R. Hofstadter and D. C. Dennett (eds). The Mind's I. Fantasies and reflections on self and soul.: New York, Bantam Books: 373-382.
Kurzweil, R. (1991). Intelligent Machines. Cambridge, The MIT Press.
Miastkowski, S. (1995). "The Penthouse Suite" Byte: 179-180.
Murray, J. H. (1997). Hamlet on the Holodeck. The future of narrative in cyberspace. New York, The Free Press.
Norman, D. A. (2004). Emotional Design. Why We Love (or Hate) Everday Things. New York, Basic Books
North, S. (1980). "Eliza. Psychoanalysis (?) by computer" in: D. Ahl and B. Green (eds) The Best of Creative Computing Volume 3.. Morristown, Creative Computing Press: 251-254, http://www.atariarchives.org
Pruijt, H. (2000). "Programming the Evolution of Cooperation." Social Science Computer Review 18(1 Spring): 77-81.
Raggett, J. and W. Bains (1992). Artificial Intelligence from A to Z. London, Chapman & Hall.
Searle, J. R. (1982). "Minds, Brains, and Programs" in: D. R. Hofstadter and D. C. Dennett (eds) The Mind's I. Fantasies and reflections on self and soul. New York, Bantam Books: 353-373.
Stephens, K. R. (2002). What Has the Loebner Contest Told Us About Conversant Systems? Cambridge Center for Behavioral Studies, http://www.behavior.org/computer-modeling/
Turing, A. M. (1950). "Computing Machinery and Intelligence." Mind. A quarterly review of psychology and philosophy LIX(236): 433-460.
Turkle, S. (1997). Life on the screen: identity in the age of the Internet. London, Phoenix.
Vickers, S. (1982). Timex Sinclair 1000 User Manual. Waterbury, Timex Computer Corporation.
Weizenbaum, J. (1966). "ELIZA - A Computer Program For the Study of Natural Language Communication between Man And Machine." Communications of the ACM 9(1): 36-45. An abridged version is available at: http://www1.uop.edu/eng/faculty/jking/ecpe151/eliza.html
Weizenbaum, J. (1984). Computer Power and Human Reason: From Judgement to Calculation. Harmondsworth, Penguin.

[1] If the input does not fit the decomposition rule, ELIZA tests the next decomposition rule associated with the keyword. If none of the decomposition rules associated with the keyword succeeds, ELIZA moves to the next keyword.

[2] This example became famous. Google reported 58.900 hits for the query +"chinese room" +Searle, October 26, 2005

[3] Hofstadter & Dennett objected that Searle overlooked the fact that it would be very hard for a human to perform the boring work of executing an AI program. (Hofstadter and Dennett 1982). Running ELIZA.XLS with tracing enabled can provide a good impression of the drudgery involved.

[4] Information in the literature about the impact of ELIZA on users is anecdotic. Fairly hard to believe, if only because it relies so much on coincidence, is the story in which someone walks up to a computer terminal, unknowingly that someone else had left the ELIZA program running on it, and then falsely assumes that it is a chat link to a real person, engages in a chat, and gets infuriated by the answers without discovering that the chat partner is a computer program. Versions of this story were presented by Kurzweil (1990: 78), who indicated that it might be a myth, and Norman (2004: 189-191) who does not express doubt about its truth value.

[5] Before the spread of personal computers, computers were mysterious machines to many. Consider the following statement from a 1982 home computer manual "You shouldn't be afraid of the computer. You are smarter than it is. So is your parakeet, for that matter." (Vickers 1982: 1).

[6] We may note that the enlightening power of Searle's "Chinese Room" example rests on the ELIZA effect. All that the example does is strip away from the Turing test the mystique of the computer, exposing the bare mechanism of rule-driven information processing.