Ideas That Created the Future

Classic Papers of Computer Science

Edited by: Harry R. Lewis

Citation:

Ideas That Created the Future: Classic Papers of Computer Science
Edited by: Harry R. Lewis
DOI: 10.7551/mitpress/12274.001.0001
ISBN (electronic): 9780262363174
Publisher: The MIT Press
Published: 2021

The MIT Press

27 ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine (1966)

Joseph Weizenbaum

Joseph Weizenbaum (1923–2008) was a German Jewish refugee who came to the United States with his family at the age of 13. After studying mathematics and computing at Wayne State University, he joined the MIT faculty in computer science. There, starting in 1964, he wrote the first example of what we would now call chatbots—programs that "know" very little but create an illusion of conversation by manipulating the discourse of their conversational partners. He was surprised that people engaged intensely with his simple program, as though it was human. Famously, his own staff assistant, who knew better than anyone that no human being was answering her musings, asked Weizenbaum to leave the room while she was using ELIZA (Weizenbaum, 1976, p. 6). It was as though she thought he was eavesdropping on a personal conversation.

ELIZA was a sensation, in part because time-sharing was new in 1966—so new that in this paper Weizenbaum felt he had to explain it to the readers of the *Communications of the ACM*. For the first time, people with no technical training were starting to use computers, and programmers started to write programs designed, in no small measure, to allow ordinary people to have some fun.

But ELIZA also touched a deep human nerve. The program is named after Eliza Doolittle, a character in George Bernard Shaw's play *Pygmalion*, which became the Broadway musical *My Fair Lady* in 1956. A film based on the musical was released in 1964. In Shaw's play, Eliza is an unschooled London flower girl who is "reprogrammed" by Professor Henry Higgins, a linguist, to feign aristocratic roots. Higgins falls in love with the (almost) perfectly transformed Eliza, in the same way that in the original Greek myth, the artist Pygmalion falls in love with the ivory statue he has sculpted of a woman.

The Greek myth of Pygmalion is in fact closer than the modern drama to the reality of ELIZA, because it involves the animation of the inanimate. In the original, Pygmalion's prayers are answered and the gods breathe life into his statue. This is only one of the Western myths of an inanimate object being brought to life in human form (see page xix).

Having witnessed as a boy the dehumanization of human beings, Weizenbaum was deeply troubled that people were so easily fooled, and skeptical of his colleagues' scientific agenda to humanize machines. He was a sharp critic of artificial intelligence throughout his life; his most important work, his attempt to separate humans and machines once and for all, was entitled *Computer Power and Human Reason* (Weizenbaum, 1976). It did not convince AI advocates,

Reprinted from Weizenbaum (1966), with permission from the Association for Computing Machinery.

and with the maturing of technologies of understanding and synthesizing speech and simulating emotion, the debate has continued about where computers *should*—as opposed to *can*—replace human interactions.

ELIZA is a program operating within the MAC time-sharing system at MIT which makes certain kinds of natural language conversation between man and computer possible. Input sentences are analyzed on the basis of decomposition rules which are triggered by key words appearing in the input text. Responses are generated by reassembly rules associated with selected decomposition rules. The fundamental technical problems with which ELIZA is concerned are: (1) the identification of key words, (2) the discovery of minimal context, (3) the choice of appropriate transformations, (4) generation of responses in the absence of key words, and (5) the provision of an editing capability for ELIZA "scripts." A discussion of some psychological issues relevant to the ELIZA approach as well as of future developments concludes the paper.

27.1 Introduction

It is said that to explain is to explain away. This maxim is nowhere so well fulfilled as in the area of computer programming, especially in what is called heuristic programming and artificial intelligence. For in those realms machines are made to behave in wondrous ways, often sufficient to dazzle even the most experienced observer. But once a particular program is unmasked, once its inner workings are explained in language sufficiently plain to induce understanding, its magic crumbles away; it stands revealed as a mere collection of procedures, each quite comprehensible. The observer says to himself "I could have written that." With that thought he moves the program in question from the shelf marked "intelligent," to that reserved for curios, fit to be discussed only with people less enlightened than he.

The object of this paper is to cause just such a re-evaluation of the program about to be "explained." Few programs ever needed it more.

27.2 ELIZA Program

ELIZA is a program which makes natural language conversation with a computer possible. Its present implementation is on the MAC time-sharing system at MIT. It is written in MAD-SLIP (Weizenbaum, 1963) for the IBM 7094. Its name was chosen to emphasize that it may be incrementally improved by its users, since its language abilities may be continually improved by a "teacher." Like the Eliza of Pygmalion fame, it can be made to appear even more civilized, the relation of appearance to reality, however, remaining in the domain of the playwright.

For the present purpose it is sufficient to characterize the MAC system as one which permits an individual to operate a full scale computer from a remotely located typewriter. The individual operator has the illusion that he is the sole user of the computer complex, while in fact others may be “time-sharing” the system with him. What is important here is that the computer can read messages typed on the typewriter and respond by writing on the same instrument. The time between the computer’s receipt of a message and the appearance of its response is a function of the program controlling the dialogue and of such MAC system parameters as the number of users currently corresponding with the system. These latter parameters generally contribute so little to the overall response time that conversational interaction with the computer need never involve truly intolerable delays.

When in conversation with ELIZA, the user types in some statement or set of statements in natural language using normal punctuation and sentence structures. Only the question mark may not be used, because it is interpreted as a line delete character by the MAC system. The user’s statement is terminated by a double carriage return which serves to turn control over to ELIZA. ELIZA then analyzes the user’s statement and generates some response which it types out. Control is then again with the user. A typical conversation is the following:

Men are all alike.
IN WHAT WAY
They’re always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE
Well, my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE
He says I’m depressed much of the time.
I AM SORRY TO HEAR YOU ARE DEPRESSED
It’s true. I am unhappy.
DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY
I need some help, that much seems certain.
WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP
Perhaps I could learn to get along with my mother.
TELL ME MORE ABOUT YOUR FAMILY
My mother takes care of me.
WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU
My father.
YOUR FATHER
You are like my father in some ways.
WHAT RESEMBLANCE DO YOU SEE
You are not very aggressive but I think you don’t want me to notice that.
WHAT MAKES YOU THINK I AM NOT VERY AGGRESSIVE
You don’t argue with me.
WHY DO YOU THINK I DON’T ARGUE WITH YOU

You are afraid of me.

DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU

My father is afraid of everybody.

WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER

Bullies.

DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE

The capitalized lines are the machine responses.

The gross procedure of the program is quite simple; the text is read and inspected for the presence of a *keyword*. If such a word is found, the sentence is transformed according to a *rule* associated with the keyword, if not a content-free remark or, under certain conditions, an earlier transformation is retrieved. The text so computed or retrieved is then printed out.

In detail, of course, the procedure sketched above is considerably more complex. Keywords, for example, may have a RANK or precedence number. The procedure is sensitive to such numbers in that it will abandon a keyword already found in the left-to-right scan of the text in favor of one having a higher rank. Also, the procedure recognizes a comma or a period as a delimiter. Whenever either one is encountered and a keyword has already been found, all subsequent text is deleted from the input message. If no key had yet been found the phrase or sentence to the left of the delimiter (as well as the delimiter itself) is deleted. As a result, only single phrases or sentences are ever transformed.

Keywords and their associated transformation rules constitute the SCRIPT for a particular class of conversation. An important property of ELIZA is that a script is data; i.e., it is not part of the program itself. Hence, ELIZA is not restricted to a particular set of recognition patterns or responses, indeed not even to any specific language. ELIZA scripts exist (at this writing) in Welsh and German as well as in English.

The fundamental technical problems with which ELIZA must be preoccupied are the following:

1. The identification of the "most important" keyword in the input message.
2. The identification of some minimal context within which the chosen keyword appears; e.g., if the keyword is "you," is it followed by the word "are" (in which case an assertion is probably being made).
3. The choice of an appropriate transformation rule and, of course, the making of the transformation itself.
4. The provision of mechanism that will permit ELIZA to respond "intelligently" when the input text contained no keywords.
5. The provision of machinery that facilitates editing, particularly extension, of the script on the script writing level.

There are, of course, the usual constraints dictated by the need to be economical in the use of computer time and storage space.

The central issue is clearly one of text manipulation, and at the heart of that issue is the concept of the *transformation rule* which has been said to be associated with certain keywords. The mechanisms subsumed under the slogan “transformation rule” are a number of SLIP functions which serve to (1) decompose a data string according to certain criteria, hence to test the string as to whether it satisfies these criteria or not, and (2) to reassemble a decomposed string according to certain assembly specifications. . . .

27.3 Discussion

At this writing, the only serious ELIZA scripts which exist are some which cause ELIZA to respond roughly as would certain psychotherapists (Rogerians). ELIZA performs best when its human correspondent is initially instructed to “talk” to it, via the typewriter of course, just as one would to a psychiatrist. This mode of conversation was chosen because the psychiatric interview is one of the few examples of categorized dyadic natural language communication in which one of the participating pair is free to assume the pose of knowing almost nothing of the real world. If, for example, one were to tell a psychiatrist “I went for a long boat ride” and he responded “Tell me about boats,” one would not assume that he knew nothing about boats, but that he had some purpose in so directing the subsequent conversation. It is important to note that this assumption is one made by the speaker. Whether it is realistic or not is an altogether separate question.

In any case, it has a crucial psychological utility in that it serves the speaker to maintain his sense of being heard and understood. The speaker further defends his impression (which may even be illusory) by attributing to his conversational partner all sorts of background knowledge, insights, and reasoning ability. But again, these are the *speaker’s* contribution to the conversation. They manifest themselves inferentially in *interpretations* he makes of the offered responses. From the purely technical programming point of view then, the psychiatric interview form of an ELIZA script has the advantage that it eliminates the need of storing *explicit* information about the real world.

The human speaker will, as has been said, contribute much to clothe ELIZA’s responses in vestments of plausibility. But he will not defend his illusion (that he is being understood) against all odds. In human conversation a speaker will make certain (perhaps generous) assumptions about his conversational partner. As long as it remains possible to interpret the latter’s responses consistently with those assumptions, the speaker’s image of his partner remains unchanged, in particular, undamaged. Responses which are difficult to so interpret may well result in an enhancement of the image of the partner, in additional rationalizations which then make more complicated interpretations of his responses reasonable.

When, however, such rationalizations become too massive and even self-contradictory, the entire image may crumble and be replaced by another (“He is not, after all, as smart as I thought he was”). When the conversational partner is a machine (the distinction between machine and program is here not useful) then the idea of *credibility* may well be substituted for that of *plausibility* in the above.

With ELIZA as the basic vehicle, experiments may be set up in which the subjects find it credible to believe that the responses which appear on his typewriter are generated by a human sitting at a similar instrument in another room. How must the script be written in order to maintain the credibility of this idea over a long period of time? How can the performance of ELIZA be systematically degraded in order to achieve controlled and predictable thresholds of credibility in the subject? What, in all this, is the role of the initial instruction to the subject? On the other hand, suppose the subject is told he is communicating with a machine. What is he led to believe about the machine as a result of his conversational experience with it? Some subjects have been very hard to convince that ELIZA (with its present script) is not human. This is a striking form of Turing's test. What experimental design would make it more nearly rigorous and airtight?

The whole issue of the credibility (to humans) of machine output demands investigation. Important decisions increasingly tend to be made in response to computer output. The ultimately responsible human interpreter of "What the machine says" is not unlike the correspondent with ELIZA, constantly faced with the need to make credibility judgments. ELIZA shows, if nothing else, how easy it is to create and maintain the illusion of understanding, hence perhaps, of judgment deserving of credibility. A certain danger lurks there.

The idea that the present ELIZA script contains no information about the real world is not entirely true. For example, the transformation rules which cause the input

Everybody hates me

to be transformed to

Can you think of anyone in particular

and other such are based on quite specific hypotheses about the world. The whole script constitutes, in a loose way, a model of certain aspects of the world. The act of writing a script is a kind of programming act and has all the advantages of programming, most particularly that it clearly shows where the programmer's understanding and command of his subject leaves off.

A large part of whatever elegance may be credited to ELIZA lies in the fact that ELIZA maintains the illusion of understanding with so little machinery. But there are bounds on the extendability of ELIZA's "understanding" power, which are a function of the ELIZA program itself and not a function of any script it may be given. The crucial test of understanding, as every teacher should know, is not the subject's ability to continue a conversation, but to draw valid conclusions from what he is being told. In order for a computer program to be able to do that, it must at least have the capacity to store selected parts of its inputs. ELIZA throws away each of its inputs, except for those few transformed by means of the MEMORY machinery. [EDITOR: A few inputs are saved in the MEMORY data structure so that things the user has mentioned earlier in the conversation can be revived when the dialog seems to have petered out.] Of course, the problem is more than one of storage. A great part of it is, in fact, subsumed under the word "selected" used just above. ELIZA in its use so far has had as one of its principal objectives the *concealment* of its lack of understanding. But to encourage its conversational partner to offer inputs from which it can select remedial information, it, must *reveal* its misunderstanding. A switch of objectives from the concealment to the revelation of misunderstanding is seen as a precondition to making an ELIZA-like program the basis for an effective natural language man–machine communication system.

One goal for an augmented ELIZA program is thus a system which already has access to a store of information about some aspects of the real world and which, by means of conversational interaction with people, can reveal both what it knows, i.e., behave as an information retrieval system, and where its knowledge ends and needs to be augmented. Hopefully the augmentation of its knowledge will also be a direct consequence of its conversational experience. It is precisely the prospect that such a program will converse with many people and learn something from each of them, which leads to the hope that it will prove an interesting and even useful conversational partner.

One way to state a slightly different intermediate goal is to say that ELIZA should be given the power to slowly build a model of the subject conversing with it. If the subject mentions that he is not married, for example, and later speaks of his wife, then ELIZA should be able make the tentative inference that he is either a widower or divorced. Of course, he could simply be confused. In the long run, ELIZA should be able to build up a belief structure (to use Abelson's phrase) of the subject and on that basis detect the subject's rationalizations, contradictions, etc. Conversations with such an ELIZA would often turn into arguments. Important steps in the realization of these goals have already been taken. Most notable among these is Abelson's and Carroll's work on simulation of belief structures (Abelson and Carroll, 1965).

The script that has formed the basis for most of this discussion happens to be one with an overwhelmingly psychological orientation. The reason for this has already been discussed. There is a danger, however, that the example will run away with what it is supposed to illustrate. It is useful to remember that the ELIZA program itself is merely a translating processor in the technical programming sense. Gorn (1964) in a paper on language systems says:

> Given a language which already possesses semantic content, then a translating processor, even if it operates only syntactically, generates corresponding expressions of another language to which we can attribute as "meanings" (possibly multiple—the translator may not be one to one) the "semantic intents" of the generating source expressions; whether we find the result consistent or useful or both is, of course, another problem. It is quite possible that by this method the same syntactic object language can be usefully assigned multiple meanings for each expression. . . .

It is striking to note how well his words fit ELIZA. The "given language" is English as is the "other," expressions of which are generated. In principle, the given language could as well be the kind of English in which "word problems" in algebra are given to high school students and the other language, a machine code allowing a particular computer to "solve" the stated problems. (See Bobrow's program STUDENT [Bobrow, 1964].)

The intent of the above remarks is to further rob ELIZA of the aura of magic to which its application to psychological subject matter has to some extent contributed. Seen in the coldest possible light, ELIZA is a translating processor in Gorn's sense; however, it is one which has been especially constructed to work well with natural language text.

This book was set in Times New Roman by the editor using LaTeX.

Library of Congress Cataloging-in-Publication Data

Names: Lewis, Harry R., editor.
Title: Ideas that created the future : classic papers of computer science /
edited by Harry R. Lewis.
Description: Cambridge, Massachusetts : The MIT Press, [2021] | Includes
bibliographical references and index.
Identifiers: LCCN 2020018950 | ISBN 9780262045308 (paperback)
Subjects: LCSH: Computer science.
Classification: LCC QA76 .I34 2020 | DDC 004–dc23
LC record available at https://lccn.loc.gov/2020018950