AI

Artificial Intelligence Fall 2023 (CSC/DSC 482.01)

Marcus Birkenkrahe

September 11, 2023

Weizenbaum ELIZA 1966

Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 9, 1 (Jan. 1966), 36–45. https://doi.org/10.1145/365153.365168

> 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 psvehiatrist "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 even in real life may 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 the *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.

What is the MAC time-sharing system? - The first modern operating system that allows users to use the same computer. The computer maintains the illusion of parallel processes by rapid task switching and prioritizing.

What were the technical problems that Weizenbaum had to solve?

1. identification of important keywords (what's being said?)
2. discovery of minimal context (what's he talking about?)
3. choice of appropriate word transformations (what to respond)
4. generation of responses in the absence of keywords (chit-chat)
5. provision of editing capability for scripts (expand conversation)

What did you think about the quality of the example conversation? - I tried to enter this conversation in modern renderings of ELIZA but I didn't get very far. I tried to feed the client statements to ChatGPT - here is the resulting conversation. Differences:

1. ChatGPT's answers are very verbose (many words).
2. ChatGPT can be told to behave like a Rogerian therapist.
3. ChatGPT can be told to "start over", and it will "forget".
4. Even as a "Rogerian therapist", ChatGPT will try to give advice.
5. Eventually, when told to "just listen", it will do that.

What is a "script" for ELIZA? - A script is a set of keywords and associated rules to transform text when a keyword is found according to the rule. A script is data, not a program, hence ELIZA is not limited to any language or a particular conversation pattern.

What do you think about subjecting language to processing? - Different views are possible: if you view language as a mechanic means of communication, nothing special, it is perfect for processing. But can

language really be separated from the human who wields it, just like the data (script) from the program in ELIZA? Perhaps the "body vs. soul" metaphor applies once again.

Does ELIZA have any memory of the conversation? - Yes, but limited to exhausting the stack of rules already applied. It throws most of its input away (it has to partly for memory reasons - in 1966).

Does ELIZA learn from experience? - Yes but not without human intervention namely by extending the script (aka keywords + rules). With that condition, it could grow ad infinitum (see CYC knowledge base).

Why does ELIZA simulate a "Rogerian therapist"? - ELIZA does not have to store explicit information about the real world, because the Rogerian therapist is supposed to direct the conversation away from himself and his knowledge (not give advice) but turn the client's questions around and help him to go deeper without suggesting content but only process. So the Rogerian therapist is an expert in exactly the type of transformation that ELIZA can do (though for different reason).

What is the importance of credibility vs. plausibility for ELIZA? - When the conversational partner is a machine then the idea of credibility ("I believe you", "It is likely") may be substituted for plausibility ("You might be right", "It is possible"). Going from 'credible' to 'plausible' accompanies the deterioration of the relationship with the user when he is convinced that his conversational partner is a machine. How often do we trust and believe? This could be a key to the success of chatbots overall.

What experimental design would make ELIZA more Turing-test capable? - Anything that makes the interface less machine-like? More of a "personality" (less servile, more dominating)?

What is meant by Bobrow's program STUDENT? - A 1964 program that could do some algebra written by Daniel G Bobrow for his PhD thesis at MIT.

What is ELIZA's principal objective? - The concealment of its lack of understanding. Does this ring true to anyone of you? In situations where you don't know really what to say or what to think?