If software is narrative: Joseph Weizenbaum, artificial intelligence and the biographies of ELIZA

If software is narrative:

Joseph Weizenbaum, Artificial Intelligence, and the Biographies of ELIZA

Author draft of manuscript accepted for publication in *New Media and Society*, September 2018

Abstract (100-150 words)

Software is usually studied in terms of the changes triggered by its operations in the material world. Yet, to understand its social and cultural impact, one needs to examine also the different narratives that circulate about it. Software's opacity, in fact, makes it prone to being translated into a plurality of narratives that help people make sense of its functioning and presence. Drawing from the case of Joseph Weizenbaum's ELIZA, widely considered the first chatbot ever created, this paper proposes a theoretical framework based on the concept of "biographies of media" to illuminate the dynamics and implications of software's discursive life. The case of ELIZA is particularly relevant in this regard because it became the center of competing narratives, whose trajectories transcended the actual functioning of this program and shaped key controversies about the implications of computing and AI.

Keywords

Artificial Intelligence, Human-machine interaction, Software, Narrative, Media imaginary, Biographies of media, History of digital media, Natural Language Processing, Joseph Weizenbaum, ELIZA

If software is narrative: Joseph Weizenbaum, Artificial Intelligence, and the Biographies of ELIZA

In July 2017, news media around the world told a worrying story. According to reports, a group of researchers at Facebook were forced to terminate an experiment with Artificial Intelligence (AI) when the conversational bot they had designed started to exchange messages with other bots using codewords that departed from everyday English (Lewis et al. 2017). Although the phenomenon had been observed before (e.g. Mordatch and Abbeel 2017) and was, therefore, not entirely unexpected, media reports presented it as a unique, troublesome discovery, linking it to prospects of a dystopian future in which machines come together to take over humans. The triviality of the episode became evident when Facebook reacted to the controversy by issuing a correction: they had shut off the project only because the company was interested in developing programs to chat with humans, and the use of incomprehensible codewords would affect their capacity to do so (McKay 2017). Lack of clarity in the press release, misunderstanding of the technical dimension, and news media's eagerness to create a sensation had turned a quite innocuous piece of software into a new occasion to fuel the moral panic about the upcoming AI apocalypse.

The controversy around Facebook's project is only one example of the dynamics by which discourses about the implications of algorithms and software shape wider debates on digital media. As Wendy Chun (2011: 2-3) aptly points out, software is inherently paradoxical, since its at times invisible character contrasts with the all-too real and visible effects it has on material reality. While many technologies are to a certain extent opaque – think, for instance, of cinema's projectors that are concealed from spectators' view during screenings (North 2007) –

software bring such opacity to a different degree. In fact, even computer scientists and programmers are often unable to reconstruct the stratifications of software and code that lead up to the actual functioning of the machine (Chun, 2011). This is even more pronounced in the case of machine learning technologies that employ neural networks, widely used in contemporary AI applications devised by companies such as Amazon, Facebook, Google and Apple: neural networks, in fact, function through complex statistical patterns whose internal functioning is hardly accessible (Burrell 2016). One of the consequences of such degree of opacity is that software is prone to be translated – sometimes even forced – into a variety of narratives that help people make sense of its functioning and presence. Because the functioning of software is often obscure, such narratives sometimes present distorted representations of its implications and impact, like in the case of Facebook's chatbot. Crucially, as a plurality of narratives are constructed and disseminated, software artefacts become contested objects whose meanings and interpretations are the subject of complex negotiations within the public sphere.

This paper moves from the recognition that software's impact should be considered at two distinct yet interrelated levels: at a material level, in terms of the changes triggered by its operations in the material world (Kitchin and Dodge 2011), and at a discursive level, in terms of the different narratives and discourses that generate and inform wider debates about technology. These two levels are interrelated because discourses about digital media have practical effects, as they contribute to inform decisions about their governance and everyday use (Mansell 2012; Crawford 2007); and conversely, because debates and discourses about software may be informed by its material uses and effects. While software's impact at a material level has been widely researched, the discursive level of software artefacts has been the subject until now of little systematic attention. This article aims to contribute to ongoing attempts (e.g. Bucher 2016) to fill this gap. This is important also because narratives about software – and about media in general – are never neutral: they contribute to shape wider visions and debates about digital technologies such as AI, with potentially meaningful repercussions on choices about the governance of these technologies (Crawford 2007). Moreover, narratives about software may also orient action: in the case of ELIZA, for instance, they contributed to set practical goals for generations of chatbot programmers (Marino 2006).

Drawing on the case study of ELIZA, widely considered the first chatbot of history, I propose a theoretical framework based on the concept of "biographies of media" to illuminate the dynamics and implications of software's discursive life. This framework allows to identify and follow throughout time the plurality of narratives through which software is presented and discussed in the public sphere. The case of ELIZA is particularly relevant in this regard not only because of its relevance in the history of digital media – the program was in fact, as Andrew Leonard (1997: 33-34) put it, "bot erectus, the first software program to impersonate a human being successfully" –, but also because ELIZA became the center of competing narratives, whose trajectories largely transcended the actual functioning of this program and shaped key controversies about the implications of computing and AI. While ELIZA's creation at the Massachusetts Institute of Technology (MIT) in 1964-66 was inspired by programmer Joseph Weizenbaum's determination to stress the illusory character of computers' intelligence, some of the narratives emerging from it reinforced the idea that machines think and understand language in similar manners as humans. Consequently, the software became a contested object that was interpreted as evidence in favor of two different, even contrasting visions: on the one side, that AI provides only the appearance of intelligence; on the other, that it may actually replicate intelligence and understanding by artificial means. In this sense, the case of ELIZA shows that the polarization of ongoing debates about AI is part of a longer history, on whose trajectory this article sheds further light.

Theoretical framework

Scholars in media studies have employed the notion of biography to illuminate the fact that media, including software, have changing material, cultural, and social dimensions upon which their trajectories are inscribed throughout time (Lesage 2013; Lesage 2016; Silverstone and Haddon 1996). My own use of the term, however, aims to stress the narrative character of biographical recounting. In this regard, biographies of media can be defined as the bodies of narratives unfolding and representing the lifespans of different media (Natale, 2016). Biographies, in fact, are a form of contingent narrative, a story that turns historical characters, places, events, and things into narratives that can be recounted and circulated through numerous channels and ways (Caine, 2010). Proposing the notion of biographies of media, in this sense, illuminates the process by which the emergence of different media and technologies throughout time is accompanied by the construction and dissemination of a range of narratives through which people make sense of these technologies and integrate them within their everyday lives.

In their most widespread sense, biographies are stories about the lifespans of individuals. Yet, these stories often entail more than this: the lives of individuals convey particular characterizations of political, ideological, and moral issues, supported by the instance of a notable or exemplary life (Furbank, 2000). In a similar way, media also become the center of numerous narratives, each of which strengthens particular visions about their implications and impact. The Web, for instance, has gone in its relatively short history through a plurality of narratives that wavered between enthusiastic visions of freedom and participation (Flichy 2007;

Mansell 2012) to dystopian views about spam, surveillance, and the “dark” web (Brunton 2013). It is impossible to understand the history of the Web without taking into account the cultural, social, and political impact of such narratives. The biographies of media approach helps contextualize and critically question narratives about media, moving beyond common trajectory arcs that alternate optimism and pessimism.

The notion of “narrative” refers here to a range of stories that are not limited to fictional narratives, but include non-fictional accounts of circumstances and events based on recurring narrative patterns or tropes. As shown in narrative theory by classical authors such as Campbell (2004) and Olney (1972), basic narrative patterns shape an extremely range variety of cultural phenomena and texts. An example of a narrative pattern that often appears in reference to media is the idea of the ‘death’ of a medium as a consequence to the introduction of a new medium: according to this recurring trope, which can be located in several moments in media history, cinema would have caused the death of print, television the death of radio and cinema, e-readers the death of print books, and so forth (Ballatore and Natale, 2016).

Approaches to storytelling in philosophy (Cavarero 2000), anthropology (Mattingly and Garro 2000), as well as media studies (Green and Brock 2000) have shown that narrative is a fundamental way to give meaning to experiences and events. As Liu (2007) aptly notes, media change stimulates a complex reactions in our social and cultural world, and the use of familiar narrative patterns helps people make sense of such changes. In other words, narrative patterns – such as the recurring trope of the “death” of a medium mentioned above – allow users to preserve the consequentiality of their everyday life against the instability of technological change (Striphas, 2009). Indeed, historians of technology have shown that technologies function not only at a material and technical level, but also through the narratives they generate or into which they are forced (Edgerton 2007; Messeri and Vertesi 2015). This is true for software as well. Think, for instance, of the Deep Blue chess program, which famously beat chess master Garry Kasparov in 1997: IBM had made considerable investments to develop Deep Blue, yet once the challenge was won and the attention of journalists evaporated, it dismantled the project and reassigned the engineers working on it to different tasks (Christian 2011: 263-64). The company was interested in the program's capacity to create a narrative – one that would symbolically place IBM at the forefront of progress in computation – more than in its potential applications beyond the Kasparov challenge.

The notion of biographies of media provides a theoretical framework as well as an analytical approach to the study of software. As a framework, it helps uncover the existence of structural patterns in the ways through which software is narrated and discussed across time. As an approach, moreover, it provides researchers with an analytical tool aimed at recognizing specific narratives across time, linking them with broader recurrent patterns in the biographies of media, and identifying the cultural and material effects that these narrative generate (see Natale 2016 for a more in-depth discussion of biographies of media's theory and approach). In applying this framework to software and to the case of Eliza, it is worth emphasizing two key characteristics of biographies of media. First, narratives about media and technologies are always plural. A medium like the Internet, for instance, has been thought of and contextualized in distinct ways in different moments and by different social actors and groups around the world. The emergence of diverse and sometimes contrasting narratives had important consequences in shaping the behavior of social groups and in orienting public discussions and policies about this medium (Streeter 2010). Second, biographies of media are often based on recurring narrative patterns that are adapted to particular objects, contexts and events. This applies also to the two narratives about ELIZA that will be discussed here. The first narrative, based on the trope of deception, recurs in a wide range of stories about the reception of new technologies throughout media history in the nineteenth and twentieth century: think, for instance, of cinema's "founding myth" entailing the story of early spectators that, according to a famous anecdote, exchanged the illusion for reality, escaping from the image of an upcoming train (Bottomore 1999). The second narrative, which presents computers as 'thinking machines', has also a long pre-history, dating back at least to the end of the seventeenth century, when comparisons between humans and machines became increasingly common (Ekbia 2008).

ELIZA, behavioral AI, and the narrative of deception

The first narrative under analysis presents ELIZA as an opportunity to expose the difference between the operations developed by computer programs and human intelligence. This narrative is epitomized by a series of anecdotes that circulated about ELIZA's reception, which revolve around the trope of deception. At a technical level, this narrative can be linked to a specific perspective within the AI field, often described as the behavioral approach. Programs designed within behavioral AI are expected to exhibit rather than to actually replicate intelligence, thereby putting aside the problem of what happens inside the machine's "brain" (Russell and Norvig 2002). For what concerns chatbots, for instance, a behavioral approach aims at producing programs capable of conducting convincing conversations, without questioning how this result is reached or if the program's cognitive processes can be compared to human intelligence. What is important is to construct the *appearance* of intelligence.

As he took up an academic position at MIT in 1964 to work on AI research, computer scientist Joseph Weizenbaum's work was informed by a similar approach (McCorduck 1979:

253). In his writings, he professed that AI was and should be distinguished from human intelligence. Yet, he made efforts to design machines that lead people into believing they were interacting with intelligent agents, since he was confident that the realization to have been deceived would help users understand the difference between human intelligence and AI (Weizenbaum 1966: 36).

Between 1964 and 1966, Weizenbaum created what is widely considered the first functional chatbot, i.e. a computer program able to interact with users via a natural language interface (Zdenek 1999: 381). The functioning of the program, called ELIZA, was rather simple. As Weizenbaum explained in the paper describing his invention, ELIZA searched the text submitted by its conversation partner for relevant keywords. When a keyword or pattern was found, the program produced an appropriate response according to specific transformation rules. These rules were based on a two-stage process by which the input was first decomposed, broking down the sentence into small segments. Then, the segments were reassembled, readapted according to appropriate rules – for instance by substituting the pronoun "you" with "I" –, and programmed words were added to produce a response. In such cases when it was impossible to recognize a keyword, the chatbot would employ preconfigured formulas, such as "I see" or "Please go on," or alternatively create a response through a "memory" structure that drew from previously inserted inputs (Weizenbaum 1966: 37; for more detailed but accessible explanations of ELIZA's functioning, see Pruijt 2006: 517-19; Wardrip-Fruin 2009: 28-32). Made available to users of the MAC time-sharing system at MIT, the program was designed to engage in conversations with human users who responded by writing on a keyboard, similar to a contemporary messaging service or online chatroom.

Weizenbaum was adamant in his contention that ELIZA exhibited not intelligence, but the illusion of it. ELIZA would demonstrate that humans in interactions with computers are vulnerable to deception (Weizenbaum 1966). As Weizenbaum conceded in an interview with historian of AI Daniel Crevier (1993: 133), ELIZA was the immediate successor of a program to play a game called Five-in-a-row or Go-MOKU, which was described in his first published paper, aptly entitled “How to Make a Computer *Appear* Intelligent” (1961, italics mine). This program used a simple strategy with no look ahead, yet it could beat anyone who played at the same naive level, and aimed at creating “a powerful illusion that the computer was intelligent” (Crevier 1993: 133). As noted in the article where it was first described, it was able to “fool some observers for some time.” Indeed, deception was, to Weizenbaum, the measure of success for the author of an AI program: “his success can be measured by the percentage of the exposed observers who have been fooled multiplied by the length of time they have failed to catch on.” On the basis of this criterion, Weizenbaum considered his program to be quite successful, as it made many observers believe that the computer behaved intelligently, providing a “wonderful illusion of spontaneity” (Weizenbaum 1961: 24).

The idea that users’ deception might be evidence of successful AI was not new. In a landmark paper anticipating many key elements of the nascent field that was to be called AI, Alan Turing had proposed in 1950 the imitation game, a practical experiment to establish if a machine can appear to humans as an intelligent agent. According to his proposal, today most commonly known as the Turing Test, human interrogators would engage in conversations with someone through a typewriter, without knowing if they were chatting with a human or a computer. If a machine succeeded in a statistically relevant way to deceive the interrogators into believing it was human, then one had to admit, if not that computers are “thinking machines,” that they could pass as such. Turing professed to believe that “in about fifty years’ time it will be possible, to programme computers (…) to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning” (Turing 1950: 443). While the implications, significance, and actual meaning of the Turing Test are still at the center of a lively debate (e.g. Warwick and Shah 2016; Levesque 2017), it is beyond doubt that the Test contributed to set practical goals among the community of AI researchers that developed in the following decades. In what is probably the most influential AI handbook in computer science, Russell and Norvig (2002: 2-3) recognize its key role in the development of the particular understanding of research in this field based on the behavioral approach, which pursues the goal of creating computers that *act* like humans.

Considering the limited computer power and resources available, ELIZA was quite successful in deceiving users, at least when the interaction was limited to a relatively brief conversation. Its efficacy was due to some intuitions that did not strictly pertain to the domain of programming, but were instead derived from insights on psychology and from Weizenbaum’s understanding of human behavior in conversations. The German-born scientist had realized that our perception of the identity of a conversation partner is crucial to the credibility of any human interaction. Thus, in order to pass convincingly for a human, a chatbot should not only respond correctly to a given input, but also play a coherent role throughout the conversation (Weizenbaum 1967). Consequently, Weizenbaum conceived ELIZA as a program that could be adapted into different roles, which he called, using one of his characteristic theatrical metaphors, *scripts*. In ELIZA’s software architecture, scripts were treated as data, which implied that they were “not part of the program itself” (Weizenbaum 1966: 37). In terms of conversation patterns, a script corresponded to a specific part that the bot would play throughout a conversation.

In the initial version of ELIZA, called DOCTOR, the program’s script simulated a psychotherapist employing the Rogerian method – a type of non-directive therapy by which the therapist reacts to the patient’s talk mainly by redirecting it back to the patient, often in the form of further questions. The choice of this role was crucial to ELIZA’s success: in fact, the dynamics of the therapy allowed the program to sustain a conversation while adding little, if anything, to it (Weizenbaum 1976a). The results are evident in some excerpts of a conversation with ELIZA that Weizenbaum published in his first paper on the subject (ELIZA’s contributions being in capital letters):

> “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. (…)” (Weizenbaum 1966: 36-37)

The choice of the name ELIZA was based, as Weizenbaum explained, from the character of Eliza Doolittle in George Bernard Shaw’s play *Pygmalion*. In the play, phonetics specialist Professor Higgins makes a bet with a friend – a scientist of linguistics himself – that he will be able with his knowledge of phonetics to train Eliza, a florist with a cockney accent, into a woman as well poised and well-spoken as a member of the aristocracy. This corresponded to Weizenbaum's hope that his program would become more refined and varied, just like the salesgirl under the tutelage of Professor Higgins (Weizenbaum 2015: 87). The choice of a literary work such as *Pygmalion*, which interrogates the issue of authenticity and performance (Shaw 1916), was not accidental. Faithful to Weizenbaum's behavioral approach, which as explained above aimed at simulating rather than replicating human intelligence, ELIZA resembled the Eliza of Pygmalion fame because it created an appearance of reality, remaining, however, "in the domain of the playwright" (Weizenbaum 1966: 36). In fact, Weizenbaum often compared his creation to an actor or actress who "improvises according to certain guidelines, that is within a certain system or - let's keep the image of the theatre - performs a certain role" (2015: 88). In order to underline that ELIZA was only able to produce the impression of reality, he went so far as to point out that "in a sense ELIZA was an actress who commanded a set of techniques but who had nothing of her own to say" and to describe it as the "parody" of a non-directive psychotherapist (Weizenbaum 1976: 188).

In the decades that followed to the creation of ELIZA, the theatrical metaphor has been used by scholars and commentators to discuss Weizenbaum's work (e.g. Murray 1998), and more broadly, it has become a common way by which commentators and developers alike describe the functioning of chatbots (Christian 2011). This is significant because, as Lakoff and Johnson (1980) famously observed, the emergence of new metaphors to describe things and events may result in orienting attitudes towards them and in guiding future actions. If the metaphor of the theatre clearly responded to Weizenbaum's willingness to demonstrate that AI is the fruit of an illusory effect, another comparison he employed points even more explicitly to the issue of deception: he noted, in fact, that users' belief that ELIZA was actually understanding what they were saying "is comparable to the conviction many people have that fortune-tellers really do have some deep insight" (Weizenbaum 1976a: 189). Like a fortune teller, ELIZA left enough space of interpretation for users to complete it, so that "the 'sense' and the continuity the person conversing with ELIZA perceives is supplied largely by the person himself" (190).

Weizenbaum was always transparent about the fact that ELIZA had limited practical application, and would be influential in shedding light on a potential path, rather than in the context of its immediate use (Loeb 2015: 17). As Margaret Boden rightly points out, in terms of programming work ELIZA was simple to the point of being obsolete even at the moment of its creation, and proved irrelevant in technical terms for the development of the field, essentially because Weizenbaum "wasn't aiming to make a computer 'understand' language" (Boden 2006: 743). His deep interest, on the contrary, on how the program would be interpreted and 'read' by users (Weizenbaum 1961; 1966; 1967) suggests that he aimed at the creation of an artefact that would produce a specific narrative about computers, AI, and the interaction between humans and machines. ELIZA, in this regard, was an artefact created to prove his point that AI is due to users' tendency to project identity – it was, therefore, a narrative about conversational programs as much as a conversational program itself. Weizenbaum expected that a particular interpretation of the program would emerge from users' interactions with the program, as they realized that the apparent intelligence of the machine was just the result of deception (Weizenbaum 1966). This narrative would present AI not as the result of humanlike intelligence programmed into the machine, but as an illusory effect. It would therefore replace the myth of the "thinking machine," which suggested that computers could equal human intelligence (Martin 1993), with a narrative more consistent with the behavioral approach in AI. By showing that a computer program of such limited complexity could trick humans into believing it were real, ELIZA would work as a demonstration of the fact that humans, facing “AI” technologies, are vulnerable to deception.

Weizenbaum’s success in turning ELIZA into the source of a specific narrative about AI as deception is evident in the stories that circulated about the software’s reception. In an anecdote that has been reported and retold numberless times, Weizenbaum told the story of his secretary who, despite being aware of how the program functioned, once asked him to leave the room, needing some privacy to chat with ELIZA.[1] Another anecdote about ELIZA concerns a computer salesman that had a teletype interchange with ELIZA, without being conscious that this was a computer program; the interaction resulted in the salesman losing his temper and reacting with fury (Boden 2006: 1352). Both anecdotes have been recalled extremely often to stress humans’ tendency to be deceived by AI’s appearance of intelligence (Wardrip-Fruin 2009) - although some, like Sherry Turkle, point out that the story of the secretary might reveal instead users’ tendency to maintain the illusion that ELIZA is intelligent “because of their own desires to breathe life into a machine” (Turkle 2005: 110).

In such anecdotes, which played a key role in informing the program’s reception, the idea that AI is the result of deception becomes substantiated through a simple, effective narrative. Indeed, one of the characteristics of anecdotes is their capacity to be remembered, retold, and disseminated, conveying meanings or claims about the person or thing they refer to. In biographies and autobiographies, for instance, anecdotes add to the narrative character of the genre, which despite being non-fictional is based on storytelling (Benton 2009), and at the same time contribute to enforce claims about the person who is the subject of the biographical sketch, e.g. her temperament, personality, and skills (Kris and Kurz 1979; Ortoleva 1996). In the reception of ELIZA, the anecdote about the secretary played a similar role, imposing a recurring pattern by which the functioning of the program was presented to the public in terms of deception. Julia Sonnevend (2016) has convincingly demonstrated that one of the characteristics of the most influential narratives about media events is their “condensation” in a single phrase and a short narrative. Wilner et al. (2014) refer to a similar process in Latourian terms as a “pasteurisation” of the narrative, by which “germs are eliminated in the name of a simple, rational and powerful explanation” (430); through the pasteurisation of the narrative, elements that do not fit with the dominant narrative about a given event are disregarded, privileging a more coherent and stable narrative. In this sense, the anecdotes about deception and particularly the story of the secretary “condensed” or “pasteurised” the story of ELIZA, presenting it to the public as a powerful narrative about computers’ capacity to deceive users.

The computer metaphor and the narrative of thinking machines

As shown above, an examination of Weizenbaum’s writings suggests that the MIT computer scientist programmed ELIZA also and specifically with the intention to present a particular vision of AI based on what can be described as the narrative of deception. This was substantiated in anecdotes about the reception of ELIZA, which circulated widely and shaped discussions about the program’s meanings and implications. Weizenbaum probably did not consider, however, that while he was able to control the program’s behavior, he would not be able to control – or to use computer science language, to ‘program’ – all the narratives emerging from it. Unexpectedly to its creator, the reception of ELIZA also entailed the emergence of a second narrative, which presented the program’s proficiency not as a successful illusion but instead as evidence that computers can rival in intelligence with humans.

Since their early history, computers have been presented as mechanical or electronic brains whose operations might be able to replicate and surpass human reason (Spufford and Uglow 1996). In narrative form, this vision corresponds to fictional stories in a wide range of science fiction scenarios where robots and computers display human characteristics (Bory and Bory 2016), as well as to journalistic reports in which computers are presented as "intelligent brains, smarter than people, unlimited, fast, mysterious, and frightening" (Martin 1993: 122). Although results in AI, especially in the early history of the field, were often far from being comparable to human intelligence, even researchers that were aware of the limits of the field tended to overstate AI's achievements, nurturing the narrative of computers as 'thinking machines' (Crevier 1993).

This narrative contrasts sharply with Weizenbaum's tenet that AI should be understood in terms of an illusory effect rather than as evidence that the machine understands and reasons like humans. Weizenbaum contended that believing that computers are thinking machines was similar to entertaining superstitious beliefs. In his paper describing ELIZA, he pointed out that computers may be regarded by laypersons as though they are performing magic. However, he argued, "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 producers, each quite comprehensible" (1966: 36). By making this point, Weizenbaum seemed unaware of a circumstance that concerns any author – from the writer of a novel to an engineer with her latest project: once their creation reaches public view, no matter how careful she or he has reflected on the meanings of its creation, these meanings can be overturned by the readings and interpretations of other writers, scientists, journalists, and laypersons. A computer program can look, despite the programmer's intention, as if performing magic; new narratives may emerge, overwriting the narrative the programmer meant to embed in the machine.

This was something Weizenbaum would learn from experience. Public reception of ELIZA, in fact, also involved the emergence of a very different narrative from the one he had intended to 'program' into the machine. With the appearance in 1968 of Stanley Kubrick's now classical science fiction film *2001: A Space Odyssey*, many thought ELIZA was "something close to the fictional HAL: a computer program intelligent enough to understand and produce arbitrary human language" (Wardrip-Fruin 2009: 32). Moreover, Weizenbaum realized that research drawing or following from his work was lead by very different understandings about the scope and goals of AI. In particular, a psychologist from Stanford University, Kenneth Mark Colby, developed a conversational bot whose design was loosely based on ELIZA, but which represented a very different interpretation of the technology. Colby hoped that chatbots would provide a practical therapeutic tool by which "several hundred patients an hour could be handled by a computer system designed for this purpose" (Colby, Watt, and Gilbert 1966). In the previous years, Weizenbaum and Colby had collaborated and engaged in discussions, and Weizenbaum expressed later some concerns that his former collaborator did not give appropriate credit to his work on ELIZA; but the main issue in the controversy that ensued between the two scientists was on moral grounds (McCorduck 1979: 313-15). A chatbot exercising therapy to real patients was in fact a prospect Weizenbaum found dehumanizing and disrespectful for patients' emotional and intellectual involvement, as he later made abundantly clear (Weizenbaum 1972; 1976a: 268-70; 2015: 81). The question arises, he contended, "do we wish to encourage people to lead their lives on the basis of patent fraud, charlatanism, and unreality? And, more importantly, do we really believe that it helps people living in our already overly machine-like world to prefer the therapy administered by machines to that given by other people?" (Weizenbaum 1976a: 269-70). This reflected the MIT computer scientist's firm belief that there are tasks that, even if theoretically or practically possible, a computer should not be programmed to do (Weizenbaum 1974).

ELIZA attracted considerable attention in the fields of computer science and AI, as well as in popular newspapers and magazines (e.g. Wilford 1968). These accounts, however, often disregarded Weizenbaum's view that its effectiveness was due to a deceptive effect. On the contrary, ELIZA's success in looking like a sentient agent was presented in a way that supported widespread narratives about computers as "thinking machines," thereby exaggerating the capabilities of AI. It was precisely what Weizenbaum had hoped to avoid since its earliest contributions to the field. This prompted him to redesign ELIZA so that the program revealed its misunderstandings, provided explanations to users, and dispelled their confusion (Weizenbaum 1967: 479; see also Geoghegan 2008: 409).

Weizenbaum worried that the ways in which the functioning of ELIZA was narrated by other scientists and in the press contributed to strengthen what he called the "computer metaphor," by which machines were compared to humans, and software's capacity to create the appearance of intelligence was exchanged for intelligence itself (Weizenbaum 1976a). The reception of ELIZA had convinced him that such dynamics of reception were "symptomatic of deeper problems" (Weizenbaum 1976a: 11): people were eager to ascribe intelligence even if there was little to warrant such a view. This could have lasting effects, he reasoned, because machines might be able to install their model of reality upon the humans who had initially built them, thereby eroding what we understand as human. As such, for Weizenbaum ELIZA demonstrated that AI was not or at least not only the technological panacea that had been enthusiastically heralded by leading researchers in the field (Dembert 1977); to him, AI awakened instead a nightmare, "the insane dream of creating a machine, a robot, that becomes human" (Weizenbaum 2015: 95).

In numerous writings over the years, Weizenbaum would lament the tension between reality and the public perception of the actual functioning of computers. The problem, he pointed out, was not only the gullibility of the public; it was also the tendency of scientists to describe their inventions in exaggerated or inaccurate ways, profiting from the fact that non-experts are unable to distinguish what might and might not be true (Weizenbaum 2015: 111).

As Noah Wardrip-Fruin notes, Eliza became in the decades after W's creation "one of the world's most famous demonstrations of the potential of computing" (2009: 24). Still today, people's tendency to believe that a chatbot is thinking and understanding like a person is often described as "Eliza effect" (Ekbia 2008: 8), describing situations in which users attribute to computer systems "intrinsic qualities and abilities which the software (…) cannot possibly achieve" (King 1995), particularly when anthropomorphization is concerned. Sherry Turkle (2005) uses this formula to describe when human users, once they knew that ELIZA or another chatbot is a computer program, change their mode of interaction so as to 'help' the bot produce comprehensible responses. In this sense, the Eliza effect would not much demonstrate the chatbot's illusory power as much as the tendency of users to fall willingly – or perhaps most aptly, complacently – into the illusion of intelligence. Yet, the phrase "Eliza effect" is usually employed with less sophistication than in Turkle's account. It has remained popular in reports of interactions with chatbots, imposing a recurrent pattern through which users' occasional inabilities to distinguish humans from computer programs is recounted (Zdenek 1999; Foner 1997).

Conclusion

Drawing on the notion of biographies of media, the analysis has shown how ELIZA became the subject of competing narratives that contributed, in such a foundational moment for the development of digital media, alternative visions about the impact and implications of computing and AI. Weizenbaum conceived the program as an artefact illustrating a theory about AI and human-computer interactions, which drew from the behavioral approach in AI to present machine's intelligence as the result of a deceptive effect. The anecdotal narratives that emerged from his endeavours, however, proved to be only part of the narratives associated with ELIZA. In fact, narratives that presented the program as an evidence of computer's intelligence – what Weizenbaum labelled the "computer metaphor" – also characterised the reception of the software. ELIZA became therefore a contested object whose different interpretations reflected and contributed to opposite visions of AI, which were destined to dominate debates about AI in the following decades and up to the present day (Gunkel 2012; Dreyfus 1972; Boden 2016).

If the narrative of the thinking machine remains one of the dominant way through which AI is discussed and presented to the public (Bory and Bory 2016), the narrative of deception continues to exercise an important role, too. This is evident, for instance, in the recent controversy about the demonstration in May 2018 of a new project by Google for an AI voice assistant, called Duplex. Similar to home assistants such as Apple's Siri or Amazon's Alexa, Google Duplex combines speech recognition, natural language processing and voice synthesizer technologies to engage in conversation with human users. Google CEO Sundar Pichai presented a recording of a conversation in which the program imitated a human voice to book an appointment with a hair salon. In order to sound more convincing and pass as human, the software included pauses and hesitation, and this strategy appeared to work as the salon representative believed to talk with a real person and accepted the reservation ("Google's AI Assistant Can Now Make Real Phone Calls" 2018). Commentaries following the demonstration stressed the importance of Google's achievements, but also the potential problems of this approach. As critical media scholar Zeynep Tufekci pointed out in a Twitter thread that circulated widely, Duplex operates "straight up, deliberate deception," opening new ethical questions concerning the capacity of users to distinguish between humans and machines.[2] Such questions promise to become more and more relevant, as studies confirm that humans tend to apply social rules and expectations to computers (Nass and Moon 2000) and as interfaces that talk and listen to users become increasingly available in computers, cars, call centres, domestic environments and toys (Nass and Brave 2005).

Looking at the narratives that characterized the reception of software artefacts such as ELIZA reminds us that the impact of software goes beyond the mere material level. Rather than being "just stories," narratives about software orient action. In the case of ELIZA, for instance, its high media profile was instrumental in orienting public discussions about AI (Boden 2016: 1353) as well as in setting practical goals for generations of programmers: a survey conducted by Mark Marino and presented in his unpublished dissertation provides evidence that communities of chatbots users and makers continue to consider ELIZA as "the progenitor, a gold standart, and the relic, a significant touchstone in current evaluations of chatbots" (Marino 2006: 8).

As argued by Zdenek, AI – but the same applies to other types of software – depends upon the production of physical as well as discursive artefacts. AI systems are "mediated by rhetoric" (Zdenek 2003: 340) because "language gives meaning, value and form" (345) to them. In order to consider the impact and implications of any piece of software, one should trace its history at both a material and a discursive level. A framework based on the notion of the biographies of media provides, in this regard, an exceedingly useful perspective to study software, especially if combined with like-minded approaches that employ the concept of biographies to unveil the changing uses and practices that concern specific pieces of software throughout time (Lesage 2016). In this regard, the notion of biographies of media has the potential to provide a theoretical framework for exploring through a diachronic perspective both the material and the discursive lives of software artefacts.

Finally, the case of ELIZA also functions as a call to consider the responsibility borne by computer scientists, reporters, science writers, and other practitioners when they contribute to the creation and dissemination of narratives about software and digital media. Some very relevant reflections about this issue can be found in Weizenbaum's writings. In a 1976 letter to the *New York Times*, he criticized the overoptimistic claims of a previously published article on self-reproducing machines, pointing to the "very special responsibility" of the science writer for the lay reader who "has no recourse but to interpret science writing very literally" (Weizenbaum 1976b: 201). If public conceptions of computing technologies are misguided, he reasoned, then public decisions about the governance of these technologies are likely to be misguided as well. As he noted in a conversation with a journalist of *The Observer* a few years later, experts' narratives could have practical effects, since "language itself becomes merely a tool for manipulating the world, no different in principle from the languages used for manipulating computers" (Davy 1982: 22). Weizenbaum's legacy, in this sense, provides us with a powerful reminder that the use of specific discourses and narratives can have practical effects, much like the instructions coded by a programmer can result in operations that trigger changes in the real world.

Acknowledgements

I would like to thank Leah Henrickson and the two anonymous reviewers for providing exceedingly useful feedback on previous versions of this article.

Works cited

Ballatore, A. & Natale, S., 2015. E-readers and the death of the book: Or, new media and the myth of the disappearing medium. *New Media & Society*, 18(10), pp.2379–2394.

Benton, Michael. 2009. *Literary Biography: An Introduction*. Malden, MA: Wiley-Blackwell.

Boden, Margaret. 2006. *Mind as Machine: A History of Cognitive Science*. *History*. Oxford: Clarendon Press. doi:10.1086/597681.

———. 2016. *AI: Its Nature and Future*. Oxford: Oxford University Press.

Bory, Stefano, and Paolo Bory. 2016. "I Nuovi Immaginari Dell'intelligenza Artificiale." *Im@go: A Journal of the Social Imaginary* 4 (6): 66–85. doi:10.7413/22818138047.

Bottomore, Stephen. 1999. "The Panicking Audience?: Early Cinema and the 'Train Effect.'" *Historical Journal of Film, Radio and Television* 19 (2). Routledge: 177–216.

Brunton, Finn. 2013. *Spam: A Shadow History of the Internet*. Cambridge, Mass.: Mit Press.

Bucher, Taina. 2016. "The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms." *Information, Communication & Society* 20 (1). Taylor & Francis: 30–44. doi:10.1080/1369118X.2016.1154086.

Burrell, Jenna. 2016. "How the Machine 'thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3 (1): 205395171562251.

doi:10.1177/2053951715622512.

Campbell, Joseph. 2004. *The Hero with a Thousand Faces*. Princeton, N.J.: Princeton University Press.

Cavarero, Adriana. 2000. *Relating Narratives: Storytelling and Selfhood*. London New York: Routledge.

Christian, Brian. 2011. *The Most Human Human: What Talking with Computers Teaches Us about What It Means to Be Alive*. London: Viking.

Chun, Wendy Hui Kyong. 2011. *Programmed Visions: Software and Memory*. Cambridge, Mass.: MIT Press.

Colby, Kenneth Mark, James P. Watt, and John P. Gilbert. 1966. "A Computer Method of Psychotherapy: Preliminary Communication." *Journal of Nervous and Mental Disease* 142 (2): 148–52.

Crawford, Susan. 2007. "Internet Think." *Journal on Telecommunications and High Technology Law* 5: 467–86.

Crevier, Daniel. 1993. *AI: The Tumultuous History of the Search for Artificial Intelligence*. New York, NY: Basic Books.

Davy, John. 1982. "The Man in the Belly of the Beast." *The Observer*, August 15.

Dembert, Lee. 1977. "Experts Argue Whether Computers Could Reason, and If They Should." *New York Times*, May 8.

Dreyfus, Hubert L. 1972. *What Computers Can't Do: A Critique of Artificial Reason*. New York: Harper & Row.

Edgerton, David. 2007. *Shock of the Old: Technology and Global History since 1900*. Oxford: Oxford University Press.

Ekbia, Hamid R. 2008. *Artificial Dreams: The Quest for Non-Biological Intelligence*. Cambridge: Cambridge University Press.

Flichy, Patrice. 2007. *The Internet Imaginaire*. Cambridge, Mass.: MIT Press.

Foner, Leonard N. 1997. “Entertaining Agents: A Sociological Case Study.” In *Proceedings of the First International Conference on Autonomous Agents*, 122–29. AGENTS ’97. New York, NY, USA: ACM. doi:10.1145/267658.267684.

Geoghegan, Bernard Dionysius. 2008. “Agents of History: Autonomous Agents and Crypto-Intelligence.” *Interaction Studies* 9 (3): 403–14.

“Google’s AI Assistant Can Now Make Real Phone Calls.” 2018. https://www.youtube.com/watch?v=JvbHu_bVa_g&time_continue=1&app=desktop.

Green, Melanie C, and Timothy C Brock. 2000. “The Role of Transportation in the Persuasiveness of Public Narratives.” *Journal of Personality and Social Psychology* 79 (5). American Psychological Association: 701.

Gunkel, David J. 2012. *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press.

King, W. 1995. “Anthropomorphic Agents: Friend, Foe, or Folly. Technical Memorandum M-95-1.”

Kitchin, Rob, and Martin Dodge. 2011. *Code/space : Software and Everyday Life*. Cambridge, Mass.: MIT Press.

Kris, Ernst, and Otto Kurz. 1979. *Legend, Myth, and Magic in the Image of the Artist: A Historical Experiment*. New Haven: Yale University Press.

Lakoff, George;, and Mark Johnson. 1980. “Metaphor We Live by.” Chicago: University of Chicago Press.

Leonard, Andrew. 1997. *Bots: The Origin of a New Species*. San Francisco: HardWired.

Lesage, Frédérik. 2013. "Cultural Biographies and Excavations of Media: Context and Process." *Journal of Broadcasting & Electronic Media* 57 (1). Routledge: 81–96. doi:10.1080/08838151.2012.761704.

———. 2016. "A Cultural Biography of Application Software." In *Advancing Media Production Research: Shifting Sites, Methods, and Politics*, edited by Chris Paterson, David Lee, and Anamik Saha, 217–32. Basingstoke, UK: Palgrave Macmillan.

Levesque, Hector J. 2017. *Common Sense, the Turing Test, and the Quest for Real AI: Reflections on Natural and Artificial Intelligence*. Cambridge, Mass.: MIT Press.

Lewis, Mike, Denis Yarats, Yann N. Dauphin, Devi Parikh, and Dhruv Batra. 2017. "Deal or No Deal? Training AI Bots to Negotiate." *Facebook Code*. https://code.facebook.com/posts/1686672014972296/deal-or-no-deal-training-ai-bots-to-negotiate/.

Loeb, Zachary. 2015. "Introduction." In *Islands in the Cyberstream: Seeking Havens of Reason in a Programmed Society*, edited by Joseph Weizenbaum, 1–31. Sacramento, CA: Liewing Books.

Mansell, Robin. 2012. *Imagining the Internet: Communication, Innovation, and Governance*. Oxford: Oxford University Press.

Marino, Mark C. 2006. "I, Chatbot: The Gender and Race Performativity of Conversational Agents." University of California, Riverside.

Martin, C Dianne. 1993. "The Myth of the Awesome Thinking Machine." *Communications of the ACM* 36 (4): 120–33.

Mattingly, Cheryl, and Linda C Garro. 2000. *Narrative and the Cultural Construction of Illness* *and Healing*. Berkeley: University of California Press.

McCorduck, Pamela. 1979. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. San Francisco, CA: W.H. Freeman.

McKay, Tom. 2017. “No, Facebook Did Not Panic and Shut Down an AI Program That Was Getting Dangerously Smart.” *Gizmodo*. https://gizmodo.com/no-facebook-did-not-panic-and-shut-down-an-ai-program-1797414922.

Messeri, Lisa, and Janet Vertesi. 2015. “The Greatest Missions Never Flown: Anticipatory Discourse and the‘ Projectory’ in Technological Communities.” *Technology and Culture* 56 (1): 54–85.

Mordatch, Igor, and Pieter Abbeel. 2017. “Emergence of Grounded Compositional Language in Multi-Agent Populations.” *arXiv Preprint arXiv:1703.04908*, 1–10.

Murray, Janet H. 1998. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Cambridge, Mass.: MIT press.

Nass, Clifford, and Scott Brave. 2005. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship.* Cambridge, Mass.: MIT press.

Nass, Clifford, and Y. Moon. 2000. “Machines and Mildlessness: Social Responses to Computers.” *Journal of Social Issues* 56 (1): 86–103.

Natale, Simone. 2016. “Unveiling the Biographies of Media: On the Role of Narratives, Anecdotes and Storytelling in the Construction of New Media’s Histories.” *Communication Theory* 26 (4): 431–449. doi:10.1111/comt.12099.

North, Dan. 2007. “Illusory Bodies: Magical Performance on Stage and Screen.” *Early Popular Visual Culture* 5 (2). Routledge: 175–88.

Olney, James. 1972. *Metaphors of Self: The Meaning of Autobiography*. Princeton, N.J.:

Princeton University Press.

Ortoleva, Peppino. 1996. “Vite Geniali: Sulle Biografie Aneddotiche Degli Inventori.” *Intersezioni* 1: 41–61.

Pruijt, Hans. 2006. “Social Interaction with Computers: An Interpretation of Weizenbaum’s ELIZA and Her Heritage.” *Social Science Computer Review* 24 (4): 516–23. doi:10.1177/0894439306287247.

Russell, Stuart J, and Peter Norvig. 2002. *Artificial Intelligence: A Modern Approach*. Upper Saddle River (NJ): Pearson Education.

Shaw, Bertrand. 1916. *Pygmalion*. New York: Brentano.

Silverstone, Roger, and Leslie Haddon. 1996. “Design and the Domestication of Information and Communication Technologies: Technical Change and Everyday Life.” In *Communication by Design: The Politics of Information and Communication Technologies*, edited by Robin Mansell and Roger Silverston, 44–74.

Sonnevend, Julia. 2016. *Stories without Borders: The Berlin Wall and the Making of a Global Iconic Event*. Oxford University Press.

Spufford, Francis, and Jennifer S Uglow. 1996. “Cultural Babbage: Technology, Time and Invention.” London: Faber.

Streeter, Thomas. 2010. *The Net Effect: Romanticism, Capitalism, and the Internet*. New York: New York University Press.

Turing, A M. 1950. “Computing Machinery and Intelligence.” *Mind* 59 (236): 433–60.

Turkle, Sherry. 2005. *The Second Self: Computers and the Human Spirit*. Cambridge, Mass.: Mit Press.

Wardrip-Fruin, Noah. 2009. *Expressive Processing: Digital Fictions, Computer Games, and*

Software Studies. Cambridge, Mass.: MIT press.

Warwick, Kevin, and Huma Shah. 2016. *Turing's Imitation Game*. Cambridge: Cambridge University Press.

Weizenbaum, Joseph. 1961. "How to Make a Computer Appear Intelligent." *Datamation* 7: 24–26.

———. 1966. "ELIZA: A Computer Program for the Study of Natural Language Communication between Man and Machine." *Communications of the ACM* 9 (1): 36–45.

———. 1967. "Contextual Understanding by Computers." *Communications of the ACM* 10 (8): 474–480.

———. 1972. "On the Impact of the Computer on Society: How Does One Insult a Machine ?" *Science* 176 (4035): 40–42.

———. 1974. "Automating Psychotherapy." *Communications of the Association for Computing Machinery* 17 (7): 425.

———. 1976a. "Computer Power and Human Reason." New York: Freeman.

———. 1976b. "Letters: Computer Capabilities." *New York Times*, March 21.

———. 2015. *Islands in the Cyberstream: Seeking Havens of Reason in a Programmed Society*. Duluth: Litwin Books.

Wilford, John Noble. 1968. "Computer Is Being Taught to Understand English." *New York Times*, June 15.

Wilner, Adriana, Tania Pereira Christopoulos, Mario Aquino Alves, and Paulo C Vaz Guimarães. 2014. "The Death of Steve Jobs: How the Media Design Fortune from Misfortune." *Culture and Organization* 20 (5). Taylor & Francis: 430–49.

Zdenek, Sean. 1999. "Rising Up from the MUD: Inscribing Gender in Software Design."

Discourse & Society 10 (3): 379–409. doi:10.1177/0957926599010003005.

———. 2003. “Artificial Intelligence as a Discursive Practice: The Case of Embodied Software Agent Systems.” *Ai & Society* 17 (3–4). Springer: 340–63.

Endnotes

[1] The gender dimension of this anecdote is also not to be overlooked, as well as the gendered identity assigned to chatbots from ELIZA to Amazon’s Alexa; see, on this, Zdenek (1999).

[2] The thread is readable at https://twitter.com/zeynep/status/994236536072953856 (accessed 28 July 2018).