Artificial Intelligence

DAVID J. GUNKEL
Northern Illinois University, USA

Whether we acknowledge it or not, communication and artificial intelligence (AI) are closely related. On the one hand, communication has been instrumental to both the theory and practice of AI. In particular, it is communication that provides the science of AI with its defining test case and experimental evidence. This is immediately apparent in the agenda-setting paper that is credited with (re)defining machine intelligence, Alan Turing's "Computing Machinery and Intelligence." Likewise the recent development of autonomous machines, learning algorithms, and smart systems introduces new challenges and opportunities for communication studies. Dealing with these innovations and their consequences will require a significant recompiling of the discipline and its traditional anthropocentric focus, instrumentalist theory of technology, and modern philosophical ideas.

The Turing test

Although the phrase "artificial intelligence" is a product of an academic conference organized by John McCarthy at Dartmouth College in 1956, it is Alan Turing's 1950 paper and its "game of imitation," or what is now routinely called "the Turing test," that defines and characterizes the field. Although Turing begins his essay by proposing to consider the question "Can machines think?" he immediately recognizes difficulties with the question itself. For this reason, he proposes to pursue an alternative line of inquiry, one that can, as he describes it, be "expressed in relatively unambiguous words."

> The new form of the problem can be described in terms of a game which we call the "imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. (Turing, 1999, p. 37)

This determination is to be made on the basis of simple questions and answers. The interrogator asks A and B various questions, and on the basis of their responses to these inquiries tries to discern whether the respondent is a man or a woman. "In order that tone of voice may not help the interrogator," Turing further stipulates, "the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms" (Turning, 1999, pp. 37–38). In this way, the initial arrangement of the "game of imitation" is, as Turing describes it,

The International Encyclopedia of Communication Theory and Philosophy.
Klaus Bruhn Jensen and Robert T. Craig (Editors-in-Chief), Jefferson D. Pooley and Eric W. Rothenbuhler (Associate Editors).

DOI: 10.1002/9781118766804.wbiect076

computer-mediated communication (CMC) *avant la lettre*. The interrogator interacts with two unknown participants via a form of synchronous computer-mediated interaction that is routinely called “chat.” Because the exchange takes place via text messages, the interrogator cannot see or otherwise perceive the identity of the two interlocutors and must, therefore, ascertain gender from responses that are supplied to questions like “Will X please tell me the length of his/her hair” (Turing, 1999, p. 37). Consequently, what Turing arranges is something that is now a familiar feature with CMC: The actual identity of one’s interlocutors is effectively hidden from view and only able to be ascertained by way of the messages that come to be exchanged.

Turing then takes his thought experiment one step further.

> We can now ask the question, “What will happen when a machine takes the part of A in this game?” Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, “Can machines think?” (Turing, 1999, p. 38)

In other words, if the man (A) in the game of imitation is replaced with a computer, would this device be able to respond to questions and “pass” as another person, effectively fooling the interrogator into thinking that it was just another human interlocutor? It is this question, according to Turing, that *replaces* the initial and unfortunately ambiguous query “Can machines think?” Consequently, if a computer does in fact become capable of successfully simulating a human being, of either gender, in communicative exchanges with a human interrogator to such an extent that the interrogator cannot tell whether he is interacting with a machine or another human being, then that machine would, Turing concludes, need to be considered “intelligent.”

At the time that Turing published the paper, he estimated that the tipping point—the point at which a machine would be able to successfully play the game of imitation—was at least half a century in the future.

> I believe that in about fifty years’ time it will be possible to programme computers, with a storage capacity of about 10^9, 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, 1999, p. 44)

It did not take that long. Already in 1966 Joseph Weizenbaum demonstrated a simple natural language processing application that was able to converse with human interrogators in such a way as to appear to be another person. ELIZA, as the application was called, was the first chatter-bot. Although this term was not utilized by Weizenbaum, it has been applied retroactively as a result of the efforts of Michael Maudlin, founder and chief scientist of Lycos, who introduced the neologism in 1994 in order to identify a similar natural language processing application he called Julia. ELIZA was, technically speaking, a rather simple program,

> consisting mainly of general methods for analyzing sentences and sentence fragments, locating so-called key words in texts, assembling sentences from fragments, and so on. It had, in other words, no built-in contextual framework or universe of discourse. This was supplied to it by a “script.” In a sense ELIZA was an actress who commanded a set of techniques but who had nothing of her own to say. (Weizenbaum, 1976, p. 188)

Despite this, Weizenbaum's program demonstrated what Turing had initially predicted:

> ELIZA created the most remarkable illusion of having understood in the minds of many people who conversed with it. People who knew very well that they were conversing with a machine soon forgot that fact, just as theatergoers, in the grip of suspended disbelief, soon forget that the action they are witnessing is not "real." This illusion was especially strong and most tenaciously clung to among people who know little or nothing about computers. They would often demand to be permitted to converse with the system in private, and would, after conversing with it for a time, insist, in spite of my explanations, that the machine really understood them. (Weizenbaum, 1976, p. 189)

Features and consequences

Although there is a good deal that has been and continues to be written in response to Turing's essay, the game of imitation, and empirical demonstrations like that provided by ELIZA and her descendants, this particular formulation of AI has four important features and consequences.

The other minds problem

Turing's essay situates communication as the deciding factor in AI. Because "the original question 'Can machines think?'" is considered by Turing to be "too meaningless," he reformulates and refers the inquiry to a demonstration of communicative ability. This is not a capricious decision; there are good philosophical reasons for proceeding in this manner, and they have to do with what philosophers routinely call "the other minds problem." "How does one determine," as Paul Churchland famously characterized it,

> whether something other than oneself—an alien creature, a sophisticated robot, a socially active computer, or even another human—is really a thinking, feeling, conscious being; rather than, for example, an unconscious automaton whose behavior arises from something other than genuine mental states? (1999, p. 67)

For Turing, as for many in the field of AI who follow his lead, intelligence is something that is neither easy to define nor able to be empirically observed. It is, therefore, evidenced in and decided on the basis of behaviors that are considered to be signs or symptoms of intelligence, especially communication in general and human-level verbal conversation in particular. In other words, because intelligent thought is not directly observable, the best one can do is deal with something, like communicative interaction, that is assumed to be the product of intelligence and can be empirically observed, measured, and evaluated.

As a consequence of this, Turing's proposal relies on and leverages a rather common assumption in the philosophy of mind, namely that communication is indicative of intelligence or at least of some kind of cognitive activity. This assumption has deep philosophical roots, going back at least to the work of René Descartes, where spoken discourse was identified as uniquely human and the only certain method by which to differentiate the rational human subject from ostensibly mindless animals and automatons. If one were, for example, confronted with a cleverly designed machine that looked and behaved like a human being, there would be, Descartes argues, at least one very certain means of recognizing that these artificial figures are in fact machines and not real men:

> They could never use words, or put together other signs, as we do in order to declare our thoughts to others. For we can certainly conceive of a machine so constructed that it utters words … But it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do. (Descartes, 1988, pp. 44–45)

Turing's game of imitation draws on and leverages this rich philosophical tradition but in reverse. If an entity—another human being, an animal, a robot, an algorithm, and so forth—is in fact capable of performing communicative operations at least on par with what is typically expected of another human individual, irrespective of what actually goes on inside the head or information processor of the entity itself, we would need to consider this entity intelligent. Following from this, Turing estimated that developments in machine communication would advance to such a degree that it would make sense to speak (and to speak intelligently) of machine intelligence by the end of the 20th century. "I predict," Turing wrote, "that by the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted" (Turing, 1999, p. 44).

Signs of intelligence versus "real intelligence"

Although this conclusion follows quite logically from Turing's argument, there has been and continues to be considerable resistance to it. For Turing, the critical challenge had been already articulated by Lady Lovelace (aka Ada Augusta Byron, the daughter of the English poet Lord Byron), who not only wrote the software for Charles Babbage's Analytical Engine but is, for that reason, considered to be the first computer scientist. "Our most detailed information of Babbage's Analytical Engine," Turing explains, "comes from a memoir by *Lady Lovelace*. In it she states, 'The Analytical Engine has no pretensions to *originate* anything. It can only do *whatever we know how to order it to perform*'" (Turing, 1999, p. 50; emphasis in the original). According to Lovelace, a computer (and at the time she wrote this, "computer" referred not to an electronic device but to a large mechanical information processor comprised of intricate gears and levers), no matter how sophisticated its programming, only does what we tell it to do. We can, in fact, write a software program like ELIZA or Apple's Siri that takes verbal input, extracts keywords, rearranges these words according to preprogrammed scripts, and then spits out seemingly intelligible results. This does not, however, necessarily mean that such a machine is capable of original thought or of understanding what is stated in even a rudimentary way.

This position is taken up and developed by John Searle in his well-known "Chinese room" example. This intriguing and influential thought experiment, introduced in 1980 with the essay "Minds, Brains, and Programs" and elaborated in subsequent publications, was offered as an argument against the claim of strong AI—that machines are able to achieve intelligent thought:

> Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing test for understanding Chinese but he does not understand a word of Chinese. (Searle, 1999, p. 115)

The point of Searle's imaginative albeit somewhat ethnocentric illustration is quite simple: simulation is not the real thing. Merely shifting verbal symbols around in a way that looks like linguistic understanding is not really an understanding of the language. Or, as Searle concludes, registering the effect of this insight on the standard test for artificial intelligence: "This shows that the Turing test fails to distinguish real mental capacities from simulations of those capacities. Simulation is not duplication" (Searle, 1999, p. 115).

Demonstrations like Searle's Chinese room, which seek to differentiate between the appearance of something and the real thing-in-itself, not only deploy an ancient philosophical distinction that is at least as old as Plato but inevitably require some kind of privileged and immediate access to the real as such and not just to how it appears. In order to distinguish, for example, between the simulation of intelligence and "real intelligence," one would need access not just to external indicators that look like intelligence but to the actual activity of intelligence as it occurs (or not) in the mind of another. This requirement, however, immediately runs into the other minds problem and the epistemological limitation that it imposes, namely, that we cannot get into the "head" of another entity—whether that entity be another human being, a nonhuman animal, an alien life form, or a machine—to know with any certitude whether they actually do or perform whatever it is they appear to manifest. In other words, Searle is only able to distinguish between and compare what appears to happen for those individuals interacting with the room and what really goes on inside the room because he has already provided privileged access to the inner workings of the room. His "counter example," therefore, violates the epistemological limitations imposed by the other minds problem, which is something Turing's game of imitation was careful to acknowledge and respect.

AI and social interaction

Even if, following Searle's argument, one is convinced that smartphones, recommendation algorithms, and computer applications remain essentially "mindless" instruments that merely manipulate linguistic tokens, the Turing test also demonstrates that it is what we do with and in response to these manipulations that makes the difference. In other words, whether we conclude that the machine is intelligent or not, the communicative behavior it exhibits in, for example, the game of imitation does have an effect on us and our social interactions and relationships. Even if one doubts the possibility of ever achieving what is called "strong AI," the fact is our world is already populated by semi-intelligent artifacts, social robots, learning algorithms, and autonomous decision-making systems that increasingly occupy the place of the Other in social relationships and communicative interaction.

This insight has been experimentally confirmed by what Clifford Nass and Byron Reeves call the "computers are social actors" (CASA) theory:

> Computers, in the way that they communicate, instruct, and take turns interacting, are *close enough* to human that they encourage *social* responses. The encouragement necessary for such a reaction need not be much. As long as there are some behaviors that suggest a social presence, people will respond accordingly … Consequently, any medium that is close enough will get human treatment, even though people know it's foolish and even though they likely will deny it afterwards. (Reeves & Nass, 1996, p. 22)

The CASA model, which was developed in response to numerous experiments with human subjects, describes how users of computers, irrespective of the actual intelligence possessed (or not) by the machine, tend to respond to the technology as another socially aware and interactive subject. In other words, even when experienced users know quite well that they are engaged with using a machine, they make what Reeves and Nass (1996, p. 22) call the "conservative error" and tend to respond in ways that afford this other thing social standing on par with another human being. Consequently, in order for something to be recognized and treated as a social actor, "it is not necessary," as Reeves and Nass (1996, p. 28) conclude, "to have artificial intelligence" strictly speaking.

This outcome is evident not only in the tightly constrained experimental studies conducted by Reeves and Nass but also in the mundane interactions with "mindless" objects like online chatter-bots and nonplayer characters, which are encountered in online communities and massively multiplayer online role playing games (MMORPGs). These software bots—ostensibly descendants of ELIZA—complicate the assumed instrumentalism operationalized by the theories of CMC, making it increasingly difficult to decide who or what is responsible for actions in the virtual space of an online community. Although these rather simple programs are by no means close to achieving anything that looks remotely like human-level intelligence or even rudimentary machine learning, they can still be mistaken for and pass as other human users. They are, in the words of Reeves and Nass, "*close enough* to human that they encourage *social* responses."

AI, technology, and communication theory

Finally, and because of this, the rules of the game in communication theory need to be significantly adjusted and modified. In the Turing test, the computer occupies the position of both medium *through* which human interlocutors exchange messages and one of the participants *with* whom one is engaged in these communicative exchanges. Despite this, communication studies has generally limited its interest and approach to the former. That is, it has typically understood and examined the computer—and all technological apparatus—as a medium of human communicative interaction. This fundamental decision concerning the role and function of the computer is consistent with the instrumental theory of technology, which understands technology as nothing more than a tool of human activity, and the standard model of communication theory, which was initially developed by Claude Shannon and Warren Weaver. Despite the remarkable success of these traditional approaches and formulations, they inevitably miss a crucial opportunity originally identified by Turing and experimentally confirmed by Reeves and Nass: the fact that a machine is not just a means of human activity and concourse but might also be a participant in communicative interactions.

Although communication studies appears to have marginalized or even ignored this other aspect, the discipline actually began by trying to address and conceptualize the full range of opportunities. This effort was initially expressed in Robert Cathcart and Gary Gumpert's 1985 essay, "The Person–Computer Interaction." In this relatively early text ("early" in terms of the discipline's recognition and engagement with computer technology), the authors draw a distinction between communicating *through* a computer and communicating *with* a computer. The former, it is argued, names all those "computer facilitated functions" where "the computer is interposed between sender and receiver." The latter designates "person–computer interpersonal functions" where "one party activates a computer which in turn responds appropriately in graphic, alphanumeric, or vocal modes establishing an ongoing sender/receiver relationship" (1985, p. 114). These two alternatives, which adhere to the configuration of Turing's game of imitation without explicitly acknowledging it, were corroborated and further refined in James Chesebro and Donald Bonsall's *Computer-Mediated Communication* (1989). In this book-length analysis, the authors detail a five-point scale that delimits the range of possibilities for "computer–human communication." The scale extends from the computer utilized as a mere medium of message transmission between human interlocutors to the computer understood as an intelligent agent with whom human users interact.

In reframing the computer according to the insights provided by this other (almost forgotten) theoretical alternative, all kinds of things change, not the least of which is our understanding of who, or what, qualifies as a legitimate subject of communication. According to Norbert Wiener, the progenitor of the science of cybernetics, these developments fundamentally alter the social landscape:

> It is the thesis of this book [*The Human Use of Human Beings*] that society can only be understood through a study of the messages and the communication facilities which belong to it; and that in the future development of these messages and communication facilities, messages between man and machines, between machines and man, and between machine and machine, are destined to play an ever-increasing part. (Wiener, 1988, p. 16)

In the social relationships of the not-too-distant future (and one needs to recall that Wiener wrote this in 1950, the same year as Turing's influential paper), the computer and related systems, like embodied robots and disembodied algorithms, will no longer be mere instruments of human communicative action or the medium through which human users communicate with each other. Instead they will occupy the position of another social actor with whom one communicates and interacts. In coming to occupy this other position, however, we inevitably run up against and encounter fundamental questions of social responsibility and ethics—questions that not only could not be articulated within the context of the previous theoretical paradigm of communication studies, but, if they had been articulated, would have been, from that perspective, considered inappropriate and even nonsense. What, for example, is our responsibility in the face of this Other—an Other who is otherwise than another human being? How do or should we respond to this other form of Otherness? How will or should this machinic Other respond to us? Although these questions appear to open onto what many would consider to be the realm of science fiction, they are already part of social reality, and the challenge for the 21st century is to decide how communication theory will respond to and accommodate these new social challenges and opportunities.

SEE ALSO: Computer-Mediated Communication

References and further readings

Cathcart, R., & Gumpert, G. (1985). The person–computer interaction: A unique source. In B. D. Ruben (Ed.), *Information and behavior* (Vol. 1, pp. 113–124). New Brunswick, NJ: Transaction.

Chesebro, J. W., & Bonsall, D. G. (1989). *Computer-mediated communication: Human relationships in a computerized world*. Tuscaloosa, AL: University of Alabama Press.

Churchland, P. (1999). *Matter and consciousness*. Cambridge, MA: MIT Press.

Descartes, R. (1637/1988). *Selected philosophical writings*, trans. J. Cottingham, R. Stoothoff, & D. Murdoch. Cambridge, UK: Cambridge University Press.

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. Cambridge, MA: MIT Press.

Partridge, D., & Wilks, Y. (1990). *The foundations of artificial intelligence*. Cambridge, UK: Cambridge University Press.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge, UK: Cambridge University Press.

Searle, J. (1984). *Minds, brains and science*. Cambridge, MA: Harvard University Press.

Searle, J. (1999). The Chinese room. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 115–116). Cambridge, MA: MIT Press.

Turing, A. (1999). Computing machinery and intelligence. In P. A. Mayer (Ed.), *Computer media and communication: A reader* (pp. 37–58). Oxford, UK: Oxford University Press.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. San Francisco, CA: Freeman.

Wiener, N. (1988). *The human use of human beings: Cybernetics and society*. Boston, MA: Da Capo Press.

David J. Gunkel is Distinguished Teaching Professor in the Department of Communication at Northern Illinois University, USA. He is the author of over 40 scholarly articles and six books: *Hacking Cyberspace*, *Thinking Otherwise: Philosophy, Communication, Technology*, *Transgression 2.0: Media, Culture and the Politics of a Digital Age*, *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*, *Heidegger and the Media*, and *Of Remixology: Ethics and Aesthetics After Remix*.