HUMAN-ROBOT INTERACTION

Do we really want AI to be human-like?

Tom Ziemke* and Sam Thellman

Behavioral variability can be used to make robots more human-like, but we propose that it may be wiser to make them less so.

Research in AI has a long tradition of making human-like behavior its goal and measuring stick. The key idea behind the Turing test (*1*) was that complex questions regarding the possibility of thought and intelligence in machines could be operationalized by testing whether people could tell if they are communicating with another person or a machine—in Turing's original version of the test interaction was reduced to text messages. Early definitions of AI often followed Turing's general strategy. Herbert Simon, for example, one of the founding fathers of AI, stated: "We call programs intelligent if they exhibit behaviors that would be regarded intelligent if they were exhibited by human beings" (*2*). In a similar vein, Rich (*3*) defined AI as "the study of how to make computers do things at which, at the moment, people are better." Although, by now, computers are better than people at many things, the idea of human-likeness as a gold standard nevertheless is still a strong influence. Writing in *Science Robotics*, Ciardo *et al.* (*4*) show that a robot equipped with human-like behavioral variability passes a nonverbal version of the Turing test, which they argue "blurs the distinction between a human and a machine." Their research makes a valuable contribution to our scientific understanding of human social cognition in general and the ascription of humanness to artifacts in particular. In this Focus, however, we wish to make the point that human-likeness is not necessarily a desirable target for the development of AI and robotic technologies.

The first thing to note is that, although in the days of early AI research, the Turing test was little more than a thought experiment, by now, several versions of the test have been developed and tested empirically. Notably, several researchers have developed nonverbal versions of the Turing test, in which people observe or interact with some robotic system. Stanton *et al.* (*5*), for example, tested whether people could distinguish the driving behavior of automated vehicles from that of human drivers. The test developed by Ciardo and colleagues involves an interactive task, in which people worked side by side with a humanoid robot that could either operate autonomously in a preprogrammed manner or be operated remotely by a person. Their experimental results indicate that participants were sensitive to humanness and correctly detected the human agent above chance level (*4*).

To put this research into context, it should be noted that variations of the Turing test can now be encountered in everyday life (*6*), as illustrated in Fig. 1. If, for example, as a pedestrian you see a car approaching the crosswalk you plan to cross (*7*), from a distance you might not be able to tell whether this is a self-driving car, so you can only judge based on the car's behavior. But even if you see somebody behind the steering wheel, at this point, you can no longer be sure whether that person is actively controlling the vehicle or maybe just monitoring its operation. Ideally, some would say, you should not even need to know whether a car is self-driving or not, because in the long run, automated cars might be better drivers than people. At the moment, however, people certainly still have mixed feelings about trusting automated vehicles. Another real-world scenario that is closer to Turing's original test arises from the fact that many companies use chatbots in their online customer service, where topics of conversation and interaction modalities are relatively limited (Fig. 1). In such settings, chatbots are often more or less indistinguishable from humans (*8*). This raises the question whether companies should disclose their chatbots' nonhuman identity, which often leads to negative consumer reactions, such as reduced trust.

As these examples illustrate, while from an engineering perspective human-like behavior might be an impressive achievement, human-machine indistinguishability raises obvious psychological, ethical, and legal questions. On the one hand, it can of course be argued that people interacting with such systems must be informed of the nature of what they are interacting with, to avoid deception (*9*). In the case of chatbots, for example, the state of California has had a chatbot disclosure law since 2018, which makes explicit disclosure a strict requirement. On the other hand, there are less clear-cut cases than chatbot versus human customer service. When it comes to automated driving, for example, interactions between a self-driving car and other road users do not have equally clear start and end points, they are typically not one-to-one, and they have certain real-time constraints. Hence, the question is exactly what should be communicated when and how, regarding a self-driving car's identity

Fig. 1. Real-world variations of the Turing test. Left: Pedestrians encountering a car at a crosswalk may not be sure to what degree it is automated. Right: A person interacting with online customer service through text messages may not have a clear picture of who or what they are interacting with.

Cognition & Interaction Lab, Human-Centered Systems Division, Department of Computer and Information Science, Linköping University, Sweden.
*Corresponding author. Email: tom.ziemke@liu.se

and capacities. Moreover, fully automated vehicles are probably still a couple of decades away; therefore, mixed traffic and varying levels of partial automation are likely to be a reality for the foreseeable future. There is much research on what external interfaces automated vehicles might need to communicate with people. However, relatively little is known about how much complexity different types of vulnerable road users (including children, people with disabilities, etc.) are actually able—and willing—to deal with. Hence, the general rule of thumb formulated above, that "people interacting with such systems must be informed of the nature of what they are interacting with," might only be possible to follow in the more clear-cut cases. This ambivalence is also reflected in discussions of social robotics research: Given the human tendency to anthropomorphize and attribute human-like mental states (*10*), many researchers aim to make robots more human-like in appearance and behavior so that they can be interacted with in more or less human-like ways. However, others would argue that robots should be easily recognizable as machines to avoid overly anthropomorphic attributions and unrealistic expectations (*7*, *9*, *10*).

In sum, as Ciardo and colleagues (*4*) point out, their work provides indications for robot design that aims at endowing robots with behavior that can be perceived by users as human-like. But, as we have argued here, it might in fact be wiser to instead use these findings to make robots less human-like. Whereas in the early days of AI, human-likeness might have been a natural target, in today's world where AI is becoming part of many people's daily life, we need to at least think twice about where it really makes sense to strive for human-like AI.

REFERENCES

1. A. M. Turing, I.—Computing machinery and intelligence. *Mind* **LIX**, 433–460 (1950).
2. H. A. Simon, C. A. Kaplan, Foundations of cognitive science, *Foundations of Cognitive Science*, M. I. Posner, Ed. (MIT Press, 1989).
3. E. Rich, *Artificial Intelligence* (McGraw-Hill, 1983).
4. F. Ciardo, D. De Tommaso, A. Wykowska, Human-like behavioral variability blurs the distinction between a human and a machine in a non-verbal Turing test. *Sci. Robot.* **7**, eabo1241 (2022).
5. N. A. Stanton, A. Eriksson, V. A. Banks, P. A. Hancock, Turing in the driver's seat: Can people distinguish between automated and manually driven vehicles? *Hum. Factors Ergon. Manuf. Serv. Ind.* **30**, 418–425 (2020).
6. S. Natale, *Deceitful Media: Artificial Intelligence And Social Life After the Turing Test* (Oxford Univ. Press, 2021).
7. T. Ziemke, Understanding robots. *Sci. Robot.* **5**, eabe2987 (2020).
8. N. Mozafari, W. H. Weiger, M. Hammerschmidt, Trust me, I'm a bot—Repercussions of chatbot disclosure in different service frontline settings. *J. Serv. Manag.* **33**, 221–245 (2022).
9. A. Sharkey, N. Sharkey, We need to talk about deception in social robotics! *Ethics Inf. Technol.* **23**, 309–316 (2021).
10. S. Thellman, M. de Graaf, T. Ziemke, *Mental State Attribution to Robots: A Systematic Review of Conceptions, Methods, and Findings* (ACM Transactions on Human-Robot Interaction, 2022); https://doi.org/10.1145/3526112.

10.1126/scirobotics.add0641

Science Robotics

Do we really want AI to be human-like?

Tom Ziemke and Sam Thellman

Sci. Robot. **7** (68), eadd0641. DOI: 10.1126/scirobotics.add0641

Science Robotics (ISSN 2470-9476) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Robotics* is a registered trademark of AAAS.