Inverting the Turing Test [review of The Most Human Human by Brian Christian]

Citation

Published Version

Permanent link

Terms of Use

Share Your Story

INVERTING THE TURING TEST

STUART M. SHIEBER

> *THE MOST HUMAN HUMAN: What Talking with Computers Teaches Us about What It Means To Be Alive.* Brian Christian. xiv + 303 pp. Doubleday, 2011. $27.95.

In his book *The Most Human Human*, Brian Christian extrapolates from his experiences at the 2009 Loebner Prize competition, a competition among chatbots (computer programs that engage in conversation with people) to see which is "most human". In doing so, he demonstrates once again that the human being may be the only animal that overinterprets.

You may not have heard of the Loebner competition, and for good reason. The annual event was inspired by the Turing test, proposed by Alan Turing in his seminal 1950 paper "Computing Machinery and Intelligence" as a method for determining in principle whether a computer possesses thought. Turing meant his test as a thought experiment to address a particular philosophical question, namely, how to define a sufficient condition for properly attributing intelligence, the capacity of thinking, to a computer. He proposed that a blind controlled test of verbal indistinguishability could serve that purpose. If a computer program were indistinguishable from people in a kind of open-ended typewritten back-and-forth, the program would have passed the test and, in Turing's view, would merit attribution of thinking.

The Loebner competition picks up on this idea; it charges a set of judges to engage in conversation with the chatbot entrants and several human confederates, and to determine which are the humans and which the computers. At the end, a prize is awarded to the "most human" chatbot—that is, the chatbot that is most highly ranked as human in paired tests against the human confederates. "Each year, the artificial intelligence (AI) community convenes for the field's most anticipated and controversial annual event," Christian says. Well, not so much. The AI community pretty much ignores this sideshow. It's the chatbot community that has taken up the Loebner competition. The Loebner prize has done little for AI beyond spreading confusions about Turing's test, some of which unfortunately find their way into Christian's (otherwise quite sound) book. For instance, Christian promulgates the common misunderstanding that Turing's test would be passed by a computer if it "fool[ed] 30 percent of human judges after five minutes of conversation." (Although

1

Turing makes a side comment about this more limited criterion, he understood the test to have no time limit and a threshold of statistical indistinguishability, which is the only philosophically sustainable stance.) And most importantly, Christian conflates the Turing test with the Loebner competition. But the two are different in many ways. In particular, on the Turing test, unlike horseshoes and hand grenades, close doesn't count. Better performance on a Turing test is not a valid basis for concluding that a machine is "closer to thinking" or "more human".

As a kind of afterthought, the promoters of the Loebner competition have been naming, in addition to the "most human" chatbot, the "most human" human—the confederate who performs best according to the same criteria in the paired tests against the chatbots. The conceit of the book is that Christian takes on a kind of moral charge to win this dubious distinction on behalf of humanity. Of course, there's nothing "most human" about the person so named. Rather, the awardee is the person whose behavior is most distinguishable from the particular bag of tricks that the chatbots happen to use. If you're interested in winning this award, it behooves you to understand how the chatbots work, and how to distinguish yourself from them.

Certainly Christian's book is successful in providing this guidance. He points out that the chatbots tend to deal with each turn in the conversation independently; they have no memory. So a good confederate will tie multiple turns of dialog together. The chatbots are light on factual knowledge, so they'll deflect questions rather than answering them. A good confederate will answer directly. This is all useful information for the confederates, but it is even more important for the judges. In reading the transcripts, you wish that the judges had read Turing's original paper, which does a great job of exemplifying a good judge's probing approach. Under that kind of attack, the chatbots would have no hope at all, which is why the Loebner competition is such thin gruel.

With so little substance, can the Loebner competition really shed any light on what makes us human? Early on, Christian talks about an only slightly tongue-in-cheek claim made by Harvard psychologist Daniel Gilbert:

> Gilbert says that every psychologist must, at some point in his or her career, write a version of "The Sentence." Specifically, The Sentence reads like this: "The human being is the only animal that _____."

When nonhumans rival humans in some area, we learn what can't fill that blank. But given the vast gulf in verbal performance between chatbots and people, they don't seem to illuminate the matter.

Why is it, then, that even with such poor performance the chatbots still hold people's interest? Why, after rehearsing all of the limitations of chatbot performance, does Christian still stick to the claim that even the primordial chatbot ELIZA's performance was "stunning, maybe even staggering"? It comes back to the issue of overinterpretation. Human cognition is geared toward finding patterns; that's what we do, and we do it well. As infants, we track the changing conditional probabilities of sounds in human speech, which allows us to learn where the boundaries are between words. As children we learn new words at a rate of several per day. We hear the patterns in music, see the objects in images, understand the logic of a story in a sequence of sentences. We can't help ourselves. We even do it when it isn't appropriate. Sure, we learn to recognize the faces of our family members and our friends. But we also see faces where there aren't any, in the moon, in clouds. We see causality where there is only coincidence—when a supplicant's prayer is answered, or when the rain dance is followed by rain. And we hear coherent language use where there are only snippets of canned text. Turing understood all this, which is why he left his Turing test wide open, unconstrained by time limit or style of questioning. Christian understands this too; the book's conceit is more honored in the breach than in the observance.

In the end, who cares about distinguishing oneself from the weak performances put on by even the best chatbots? I expect not even Christian, who admits the award is "disappointing, anticlimactic." The book is most successful when it roams farthest from its ostensible subject. Christian uses the hook of pursuing the "most human human" award as an opportunity to explore a wide range of matters related, more or less, to issues of language use by humans and computers, and it is here that the book is most rewarding and entertaining. He riffs from the statelessness of chatbot conversation to Markov chains to information theory to text compression to morality, from phonagnosia to speed dating to the coherence of personality to the television show *The Office* to the travails of customer service. The connections of these subjects to the hook and to one another are tangential at best, but that's okay in a popular science book. These are fascinating topics more or less related to cognitive science and its broad connections to the humanities. The "most human human" bit is the MacGuffin.

Stuart M. Shieber is James O. Welch, Jr., and Virginia B. Welch Professor of Computer Science and director of the Office for Scholarly Communication at Harvard University. His most recent book is *The Turing Test* (MIT Press, 2004).