

COMPUTER SCIENCE

Beyond the Turing Test

Concluding that there is no one test for machine intelligence, AI researchers develop a battery of research challenges

By Jia You

As the movie *The Imitation Game* celebrates British mathematician Alan Turing's contributions to the Allied victory in World War II, the artificial intelligence (AI) community is rethinking another of his legacies: the Turing Test.

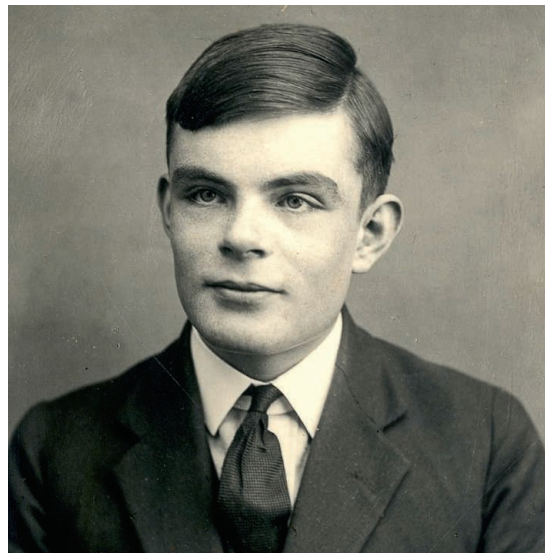
In 1950, Turing laid out an appealingly simple test for whether a machine possesses human-level intelligence: Will a person conversing with it in text mistake it for another human being? But more than 60 years later, it's time for new criteria, says computer scientist Leora Morgenstern of Leidos Inc. in Reston, Virginia. "We now know a lot about AI and what's needed to make progress. It's a big leap from Turing's time."

At a 25 January workshop at the 29th Association for the Advancement of Artificial Intelligence conference in Austin, Morgenstern and other researchers will discuss proposals for a new Turing Championship. In contrast with Turing's single litmus test, the proposed challenges acknowledge that intelligence has multiple dimensions—from language comprehension to social awareness—that are best tackled piece by piece.

Over the years, Turing's original idea has grown into a small industry while drawing increasing criticism. Competitions such as the long-running Loebner Prize ask human judges to text chat with either a person or a computer program for less than 30 minutes and then determine the converser's identity. In June, a computer program named Eugene Goostman, which adopts the persona of a 13-year-old Ukrainian boy, was declared to have passed a Turing Test organized by the University of Reading in the United Kingdom after fooling a third of the judges in 5-minute conversations. Yet researchers such as cognitive scientist Gary Marcus of New York University in New York City argue that such competitions put a premium on stock answers and other ruses. "It's a parlor trick," Marcus says. "There's no sense in which that program is genuinely intelligent." The new Turing Championship

would motivate researchers to develop machines with a deeper understanding of the world, argues Marcus, who is co-organizing the workshop.

One set of proposed challenges focuses on common-sense reasoning, which remains a tall order for machines yet is crucial for comprehending language. Take the sentence, "The trophy would not fit in the



"We can only see a short distance ahead, but we can see plenty there that needs to be done."

Alan Turing, 1950

brown suitcase because it was too big." Deducing that "it" refers to the trophy, not the suitcase, requires general knowledge that is second nature for a person but difficult to program into a machine. Next fall, in what could be the first of the new Turing challenges, the industry-sponsored Winograd Schema Challenge will test machines' comprehension of such grammatically ambiguous sentences.

A second set of proposed challenges centers on machine vision. With new machine-learning techniques that train computers to discern objects, researchers at places such as Google and Facebook are developing algorithms that can guide a self-driving car or automatically identify any face in any

photograph. But AI researchers want machines to understand and reason with what they see, says computer scientist Fei-Fei Li of Stanford University in Palo Alto, California. The challenge Li will propose would ask machines to tell stories from pictures—not only identifying an object such as a coffee mug, for example, but also noting that it sits half-empty on a table because someone drank from it. Such machines might one day interpret what she calls the "dark matter of the digital age": images and videos, which today's search engines and bots can hardly make sense of.

For machines to truly assist people in their daily lives, physical movement smoothly integrated with language and perceptual skills has to be part of the mix, says computer scientist Charles Ortiz of the Nuance Natural Language and AI Laboratory in Sunnyvale, California. His proposed challenge would ask both a machine and a human to manipulate a robotic arm in order to, say, play with a toy. At the same time, they would carry on a conversation about their actions. As in Turing's original test, a judge would evaluate the "humanness" of the computer's performance.

Intelligence has one more dimension, says computer scientist Barbara Grosz of Harvard University: teamwork. To effectively collaborate with humans, machines will need to understand their teammates' preferences, share information appropriately, and handle uncertain environments. Grosz's challenge would pair computers with people in group activities, such as formulating health care plans, to test whether people overlook that their partners aren't human.

Many more research challenges will be debated at the workshop, aimed at capabilities from long-term learning to creativity. The goal, Marcus says, is to winnow the proposals down to three to five competitions. A balance of ambition and realism is key, says computer scientist Stuart Shieber of Harvard. "You want to design competitions that are qualitatively beyond the current level of AI, but not so far that ... it would be like setting an X prize for space flight in da Vinci's era," he says.

Although it's unlikely that consensus will emerge in January, the discussion will continue at another AI conference in July, says co-organizer Manuela Veloso of Carnegie Mellon University in Pittsburgh, Pennsylvania. By early 2016, the organizers hope to stage a set of trial competitions that will be revised and repeated regularly. "If we don't move fast, it won't happen," Veloso says. "People will lose momentum." ■