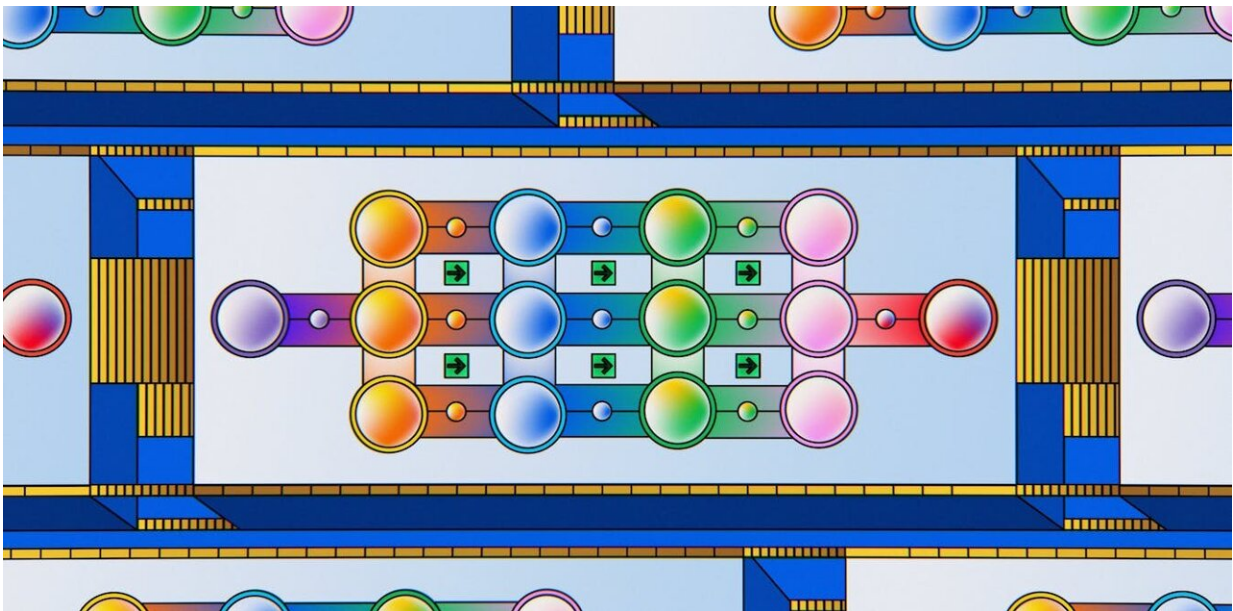# AI is closer than ever to passing the Turing test for 'intelligence'. What happens when it does?

October 17 2023, by Simon Goldstein and Cameron Domenico Kirk-Giannini

Credit: Pexels/Google Deepmind, CC BY-SA

In 1950, British computer scientist Alan Turing proposed an experimental method for answering the question: can machines think? He suggested if a human couldn't tell whether they were speaking to an artificially intelligent (AI) machine or another human after five minutes of questioning, this would demonstrate AI has human-like intelligence.

Although AI systems remained far from passing Turing's test during his lifetime, he speculated that "[…] in about 50 years' time it will be possible to program computers […] to make them play the imitation game so well that an average interrogator will not have more than 70% chance of making the right identification after five minutes of questioning."

Today, more than 70 years after Turing's proposal, no AI has managed to successfully pass the test by fulfilling the specific conditions he outlined. Nonetheless, as [some headlines](#) [reflect](#), a few systems have come quite close.

[One recent experiment](#) tested three large language models, including GPT-4 (the AI technology behind ChatGPT). The participants spent two minutes chatting with either another person or an AI system. The AI was prompted to make small spelling mistakes—and quit if the tester became too aggressive.

With this prompting, the AI did a good job of fooling the testers. When paired with an AI bot, testers could only correctly guess whether they were talking to an AI system 60% of the time.

Given the rapid progress achieved in the design of natural language processing systems, we may see AI pass Turing's original test within the next few years.

But is imitating humans really an effective test for intelligence? And if not, what are some alternative benchmarks we might use to measure AI's capabilities?

Limitations of the Turing test

While a system passing the Turing test gives us "some" evidence it is

intelligent, this test is not a decisive test of intelligence. One problem is it can produce "false negatives."

Today's large language models are often designed to immediately declare they are not human. For example, when you ask ChatGPT a question, it often prefaces its answer with the phrase "as an AI language model." Even if AI systems have the underlying ability to pass the Turing test, this kind of programming would override that ability.

The test also risks certain kinds of "false positives." As philosopher Ned Block pointed out in a 1981 article, a system could conceivably pass the Turing test simply by being hard-coded with a human-like response to any possible input.

Beyond that, the Turing test focuses on human cognition in particular. If AI cognition differs from human cognition, an expert interrogator will be able to find some task where AIs and humans differ in performance.

Regarding this problem, Turing wrote, "This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection."

In other words, while passing the Turing test is good evidence a system is intelligent, failing it is not good evidence a system is "not" intelligent.

Moreover, the test is not a good measure of whether AIs are conscious, whether they can feel pain and pleasure, or whether they have moral significance. According to many cognitive scientists, consciousness involves a particular cluster of mental abilities, including having a working memory, higher-order thoughts, and the ability to perceive one's environment and model how one's body moves around it.

The Turing test does not answer the question of whether or not AI systems have these abilities.

AI's growing capabilities

The Turing test is based on a certain logic. That is: humans are intelligent, so anything that can effectively imitate humans is likely to be intelligent.

But this idea doesn't tell us anything about the nature of intelligence. A different way to measure AI's intelligence involves thinking more critically about what intelligence is.

There is currently no single test that can authoritatively measure artificial or human intelligence.

At the broadest level, we can think of intelligence as the ability to achieve a range of goals in different environments. More intelligent systems are those which can achieve a wider range of goals in a wider range of environments.

As such, the best way to keep track of advances in the design of general-purpose AI systems is to assess their performance across a variety of tasks. Machine learning researchers have developed a range of benchmarks that do this.

For example, GPT-4 was able to correctly answer 86% of questions in massive multitask language understanding—a benchmark measuring performance on multiple choice tests across a range of college-level academic subjects.

It also scored favorably in AgentBench, a tool that can measure a large language model's ability to behave as an agent by, for example, browsing

the web, buying products online and competing in games.

Is the Turing test still relevant?

The Turing test is a measure of imitation—of AI's ability to simulate the human behavior. Large language models are expert imitators, which is now being reflected in their potential to pass the Turing test. But intelligence is not the same as imitation.

There are as many types of intelligence as there are goals to achieve. The best way to understand AI's intelligence is to monitor its progress in developing a range of important capabilities.

At the same time, it's important we don't keep "changing the goalposts" when it comes to the question of whether AI is intelligent. Since AI's capabilities are rapidly improving, critics of the idea of AI intelligence are constantly finding new tasks AI systems may struggle to complete—only to find they have jumped over [yet another hurdle](#).

In this setting, the relevant question isn't whether AI systems are intelligent—but more precisely, what "kinds" of [intelligence](#) they may have.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation