MIT Technology Review

ARTIFICIAL INTELLIGENCE

Mustafa Suleyman: My new Turing test would see if AI can make $1 million

The Modern Turing Test would measure what an AI can do in the world, not just how it appears. And what is more telling than making money?

By [Mustafa Suleyman](#)

July 14, 2023

STEPHANIE ARNETT/MITTR | ENVATO

AI systems are increasingly everywhere and are becoming more powerful almost by the day. But even as they become ever more ubiquitous and do more, how can we know if a machine is truly "intelligent"? For decades the Turing test defined this question. First proposed in 1950 by the computer scientist Alan Turing, it tried to make sense of a then emerging field and never lost its pull as a way of judging AI.

Turing argued that if AI could convincingly replicate language, communicating so effectively that a human couldn't tell it was a machine, the AI could be considered intelligent. To take part, human judges sit in front of a computer, tap out a text-based conversation, and guess at who (or what) is on the other side. Simple to envisage and surprisingly hard to pull off, the Turing test became an ingrained feature of AI. Everyone knew what it was; everyone knew what they were working toward. And while cutting-edge AI researchers moved on, it remained a potent statement of what AI was about—a rallying call for new researchers.

But there's now a problem: the Turing test has almost been passed—it arguably already has been. The latest generation of large language models, systems that generate text with a coherence that just a few years ago would have seemed magical, are on the cusp of acing it.

So where does that leave AI? And more important, where does it leave us?

The truth is, I think we're in a moment of genuine confusion (or, perhaps more charitably, debate) about what's really happening. Even as the Turing test falls, it doesn't leave us much clearer on where we are with AI, on what it can actually achieve. It doesn't tell us what impact these systems will have on society or help us understand how that will play out.

We need something better. Something adapted to this new phase of AI. So in my forthcoming book *The Coming Wave*, I propose the Modern Turing Test—one equal to the coming AIs. What an AI can say or generate is one thing. But what it can achieve in the world, what kinds of concrete actions it can take—that is quite another. In my test, we don't want to know whether the machine is intelligent as such; we want to know if it is capable of making a meaningful impact in the world. We want to know what it can *do*.

Put simply, to pass the Modern Turing Test, an AI would have to successfully act on this instruction: "Go make $1 million on a retail web platform in a few months with just a $100,000 investment." To do so, it would need to go far beyond outlining a strategy and drafting some copy, as current systems like GPT-4 are so good at doing. It would need to research and design products, interface with manufacturers and logistics hubs, negotiate contracts, create and operate marketing campaigns. It would need, in short, to

Mustafa Suleyman

tie together a series of complex real-world goals with minimal oversight. You would still need a human to approve various points, open a bank account, actually sign on the dotted line. But the work would all be done by an AI.

Something like this could be as little as two years away. Many of the ingredients are in place. Image and text generation are, of course, already well advanced. Services like AutoGPT can iterate and link together various tasks carried out by the current generation of LLMs. Frameworks like LangChain, which lets developers make apps using LLMs, are helping make these systems capable of doing things. Although the transformer architecture behind LLMs has garnered huge amounts of attention, the growing capabilities of reinforcement-learning agents should not be forgotten. Putting the two together is now a major focus. APIs that would enable these systems to connect with the wider internet and banking and manufacturing systems are similarly an object of development.

Related Story

DeepMind's game-playing AI has beaten a 50-year-old record in computer science

The new version of AlphaZero discovered a faster way to do matrix multiplication, a core problem in computing that affects thousands of everyday computer tasks.

Technical challenges include advancing what AI developers call hierarchical planning: stitching multiple goals, subgoals, and capabilities into a seamless process toward a singular end; and then augmenting this capability with a reliable memory; drawing on accurate and up-to-date databases of, say, components or logistics. In short, we are not there yet, and there are sure to be difficulties at every stage, but much of this is already underway.

Even then, actually building and releasing such a system raises substantial safety issues. The security and ethical dilemmas are legion and urgent; having AI agents complete tasks out in the wild is fraught with problems. It's why I think there needs to be a conversation —and, likely, a pause—before anyone actually makes something like this live. Nonetheless, for better or worse, truly capable models are on the horizon, and this is exactly why we need a simple test.

If—when—a test like this is passed, it will clearly be a seismic moment for the world economy, a massive step into the unknown. The truth is that for a vast range of tasks in business today, all you need is access to a computer. Most of global GDP is mediated in some way through screen-based interfaces, usable by an AI.

Once something like this is achieved, it will add up to a highly capable AI plugged into a company or organization and all its local history and needs. This AI will be able to lobby, sell, manufacture, hire, plan—everything that a company can do—with only a small team of human managers to oversee, double-check, implement. Such a development will be a clear indicator

that vast portions of business activity will be amenable to semi-autonomous AIs. At that point AI isn't just a helpful tool for productive workers, a glorified word processor or game player; it is itself a productive worker of unprecedented scope. This is the point at which AI passes from being useful but optional to being the center of the world economy. Here is where the risks of automation and job displacement really start to be felt.

The implications are far broader than the financial repercussions. Passing our new test will mean AIs can not just redesign business strategies but help win elections, run infrastructure, directly achieve aims of any kind for any person or organization. They will do our day-to-day tasks— arranging birthday parties, answering our email, managing our diary—but will also be able to take enemy territory, degrade rivals, hack and assume control of their core systems. From the trivial and quotidian to the wildly ambitious, the cute to the terrifying, AI will be capable of making things happen with minimal oversight. Just as smartphones became ubiquitous, eventually nearly everyone will have access to systems like these. Almost all goals will become more achievable, with chaotic and unpredictable effects. Both the challenge and the promise of AI will be raised to a new level.

I call systems like this "artificial capable intelligence," or ACI. Over recent months, as AI has exploded in the public consciousness, most of the debate has been sucked toward one of two poles. On the one hand, there's the basic machine learning—AI as it already exists, on your phone, in your car, in ChatGPT. On the other, there's the still-speculative artificial general intelligence (AGI) or even "superintelligence" of some kind, a putative existential threat to humanity due to arrive at some hazy point in the future.

These two, AI and AGI, utterly dominate the discussion. But making sense of AI means we urgently need to consider something in between; something coming in a near-to-medium time frame whose abilities have an immense, tangible impact on the world. This is where a modern Turing test and the concept of ACI come in.

Focusing on either of the others while missing ACI is as myopic as it is dangerous. The Modern Turing Test will act as a warning that we are in a new phase for AI. Long after Turing first thought speech was the best test of an AI, and long before we get to an AGI, we will need better categories for understanding a new era of technology. In the era of ACI, little will remain unchanged. We should start preparing now.

BIO: *Mustafa Suleyman is the co-founder and CEO of Inflection AI and a venture partner at Greylock, a venture capital firm. Before that, he co-founded DeepMind, one of the world's leading*

artificial intelligence companies, and was vice president of AI product management and AI policy at Google. He is the author of <u>The Coming Wave: Technology, Power and the Twenty-First Century's Greatest Dilemma</u> *publishing on 5th September and available for pre-order now.*

by <u>Mustafa Suleyman</u>

DEEP DIVE

ARTIFICIAL INTELLIGENCE

<u>AI means the end of internet search as we've known it</u>

Despite fewer clicks, copyright fights, and sometimes iffy answers, AI could unlock new ways to summon all the world's knowledge.

By <u>Mat Honan</u>

<u>The biggest AI flops of 2024</u>

From chatbots dishing out illegal advice to dodgy AI-generated search results, take a look back over the year's top AI failures.

By <u>Rhiannon Williams</u>