Turing Test 2.0: A Better Way to Test Machine Intelligence?

https://mindmatters.ai/podcast/ep360

Robert J. Marks:

Greetings and welcome to Mind Matters News. I'm your Turing testable host, Robert J. Marks. The Turing Test proposed by Alan Turing in 1950 is a method for assessing a machine's intelligence by evaluating whether it can imitate human conversation so convincingly that a human judge cannot reliably distinguish it from another human being. In the test an interrogator communicates with both a human and a machine usually by text, so the voice doesn't give it away and tries to identify which is which. If the judge cannot tell them apart, the machine to is said to have demonstrated intelligence. Our guest today is Georgios Mappouras who says the Turing Test is not enough to measure intelligence. And I agree with them. There are some AI researchers that they think that while modern AI can simulate human-like conversation impressively, this does not mean it is actually passed the Turing Test in a rigorous, sustained and meaningful way.

I don't agree with this. I think the Turing Test has been passed by large-language models already, such as Grok and ChatGPT, at least on a rudimentary basis. I think the Lovelace Test first proposed by summer Bringsjord is a good test of creativity and creativity being a component of intelligence. Our guest today proposes the Turing Test 2.0 that it's a more rigorous testing of intelligence of AI, the shortcomings of the original AI, its reliance on things such as deception and imitation and shallowness are one reason new proposals aim at setting clearer standards for detecting true intelligence. A link to his paper entitled, The General Intelligence Threshold, is provided in the podcast notes. Our guest today is Dr. Georgios Mappouras and he's the one that is proposing this Turing Test 2.0. I have his permission to call him George. So that's what we're going to do for the remainder of the podcast.

And George, you're welcome to call me Bob. He was born in Greece. He grew up in Cyprus and he is a study or a graduate in electrical and computer engineering at the National Technical University of Athens, which of course is in Greece. And after graduation in 2014, he moved to the US to pursue a PhD in computer architecture at Duke University. Really great school. He received his PhD about a quarter-century ago, and he is currently working in Silicon Valley, including Oracle. And this is where technically gifted people go to get high paying jobs and maybe to get rich. George, welcome.

Georgios Mappouras:

Hi Bob. Thank you for having me.

Robert J. Marks:

I want to get something out of the way first, not to do with the topic, but we were talking before about what it's like to live in Silicon Valley. Everybody's heard of Silicon Valley, but they don't know very much about it. First of all, I understand a beautiful place to live. Is that right?

Georgios Mappouras:

Yeah, especially for me coming from Cyprus, being used to the Mediterranean climate, the nice weather, sunny every day. Yeah, this is the perfect place, you want that type of weather.

Robert J. Marks:

What about it culturally? I think you mentioned there's lots of venues there that you can go to and it's really rich and it's a good place to get employment. If you work for one company and you decide maybe to go to another company, it's easy to hop ship. Is that right?

Georgios Mappouras:

Yeah. So I don't want to make it sound like a paradise. Every place has its own drawbacks. Yeah, this is the thing with Bay Area because of the companies and because a lot of employees come from different places, say an international element to it. So in terms of food or entertainment, it feels like you can find many different cultures. So especially if you're not an American, I guess that you can find things that you cannot as easily access in other places in the US.

Robert J. Marks:

Okay. Is there a good Greek community there?

Georgios Mappouras:

Not big, because I guess we're not that many, but... Yeah, a good one. Yeah.

Robert J. Marks:

Okay. Now, I was telling George before we started recording that I spent many years in Seattle and I think Seattle is a beautiful place. I wouldn't care to live there today, but in the summer it's the one of the most beautiful places in the world. And the joke in Seattle is during the winter, it's like living in a car wash. It's terrible. Now George rolled his eyes. You think that the climate is better in the Bay in Silicon Valley?

Georgios Mappouras:

Yeah, of course. Yeah. So I've visited Seattle, I visited other places in the US too, in the North. Yeah, I don't think I could permanently live there. Yeah, I don't adjust well to cloudy weather or cold weather. Yeah.

Robert J. Marks:

Oh, yeah, yeah.

Georgios Mappouras:

I can visit. I like visiting, but... Yeah.

Robert J. Marks:

It's actually terrible in the winter. It's just gloomy. In fact, I have my son and his wife and two of my grandkids live in the Seattle area, and one of the things that happens is that the sun doesn't shine, they don't get their vitamin D and they become depressed and they literally have to buy sun lamps and sit under sun lamps, and I believe take vitamin D in order to get rid of their depression. It's called SADS, and I forget what the anachronism means, but it's this idea that during the gloomy part of the year that you kind of get depressed. It's rough. So I don't think you have that in San Francisco.

Georgios Mappouras:

No, no, definitely not. Yeah.

Robert J. Marks:

Okay. Have you ever seen Alcatraz, by the way? Do you drive by Alcatraz ever?

Georgios Mappouras:

I live in the Bay Area, but I visit San Francisco sometimes. And if you drive by the coast, yeah, you can see it. I haven't visited, but you can see. I know there is... I don't know if call them cruise or boats that take you there to see the place, but I haven't done that.

Robert J. Marks:

Okay. Yeah, I would always like to visit Alcatraz. I've seen it in the movies and it's very famous. So let's get down to business. Let's talk about your paper. You wrote a paper, Turing Test 2.O, the General Intelligence Threshold. And this is a new and I think innovative measure of to whether AI becomes intelligent. You and I don't agree on everything, but I think that this is really an innovative paper. So let me ask you this, what motivated you to rethink the Turing Test?

Georgios Mappouras:

Yeah, so the first thing that started me asking the question was probably hearing people talk about AI and let's say describing its abilities without providing evidence. It was more like, I feel like this is going to happen. So I was like, "I want to see, I want to actually find a way to test it." So the first thing is like, "Okay, let's see how people test AI right now, how intelligent is it?" And looking at the Turing Test, the first thing that made me, let's say, question it a little bit was that we have this interrogation system and we have a human in a machine. I'm like, "Okay, how do I select a human that's any human suffice?" Because not all humans have the same intelligence.

So do I get an average? Do I get, let's say, am I going to question you about physics and I get an expert in physics? And then what is the line of question? Because different people will come with different lines of questioning. And let's say maybe for some humans the machine deceives me, for some does not. What if I take it even? What if I think the human is the machine? What does that mean, you see? So it felt like it had a lot of inconclusive, let's say, outcomes, which made me feel like ambiguous and not well defined. So this is the first things that made me feel like this is not enough.

Robert J. Marks:

Yeah, that's a good point. By the way, I have a reverse Turing Test. If I were to be in a conversation and wondered whether it was a human or computer, I would ask the computer, "What's the cube root of 2,416?" And if it gave me an immediate answer, I would know it was a computer. So your point is well taken. If a physicist talks to it and he says, "What's the equation for, I don't know, Newton's laws," or something like that, that's different than from a kindergartner talking to it and asking it about ... I don't know....

Georgios Mappouras:

Exactly. And it doesn't really mean that it's not intelligent, you can tell it's a machine, but that does not mean it's not intelligent because of that test. You see what I mean? If you can do intelligent things, but I can still tell it's a machine because of how it talks. And this is actually my second point that I started thinking about is we think about it through the natural language. I either speak to the machine and my question was why? Yeah, sure, humans do use natural language, but what about other types of communication? What if I tell lead to draw pictures and have two test subjects? They both draw picture.

Why do we use the language? It is like these things feels like selected without a rigorous process. We just said, "Okay, that's how humans do it. Let's just do that." And I think Turing Test doesn't test if it's intelligent, test how good you can mimic a human, but not all human attributes are due to intelligence.

Robert J. Marks:

Yeah, that's true. Yesterday I gave a talk before the Waco Chamber of Commerce and I was joined by the CEO of a company called Worlds, and they do AI where they do their learning from images, I guess that most of the fuel for training large language models has been exhausted. And so this guy says, "There's all sorts of information in the world if you just know how to extract it from photos and things of that sort." And one of the things he did, George, which I found fascinating, and I'm going to start to do it, I usually when I drive or commute, I listened to podcasts.

He said he doesn't do that anymore. What he does is he puts on Grok and has a conversation with it and he asks different topical questions and they go back and forth and he says he learns a lot. He's become an expert in a lot of areas. So I want to start doing that. I thought that's a good substitution for podcasts. So that's what I want to do in the future. One of the big things today is this idea of artificial general intelligence, AGI, and I think that the definition varies from place to place. Do you have a definition of AGI, artificial general intelligence that you go by?

Georgios Mappouras:

So that's one of the things, the other things I didn't like in this, let's say, research area, let's call it, is that everybody comes with their own definition and then we don't really know if everybody has their own definition, how do we look for it? So first of all, I like to drop the artificial, I don't care how it's generated.

Robert J. Marks:

Oh, that's interesting. Okay.

Georgios Mappouras:

So if it's general intelligence, even general intelligence, sometimes it doesn't, but let's stick with it. Is it truly intelligent? That's what we really want to know. And let's say we call that the human level of intelligence, but whatever that means. So if it's general intelligence or artificial general intelligence, if they reach the same level, let's say, it's the same thing. So we don't need two terms. So it's general intelligence. So this is the first thing that I noticed that it doesn't matter how it's generated maybe from an alien species, maybe from a machine, maybe from a human, is that level. So now the important thing is to define that level. So let's call it GI for sure, general intelligence. And that's the first thing that I tried to do, try to pinpoint it. And in order to pinpoint it, I actually looked at the human race as a system.

So we can look at the human race as a system and think about what does it do, what does this thing that it does that we actually find is fascinating? And we can look, for example, compare this system with other systems, let's say, another animal like a dog. And we can see how the dog from its, let's say, from the time it started its life on the earth up to this point, how it lives. And we can look at humans and how they live, and we can see that humans have a unique thing, at least from the data we can get now we see that humans started living, let's say, from caves putting up fire. Now we live in skyscrapers, we travel to the moon.

So what we see is that there's an increase in knowledge, there's an increase in knowledge that's increasing information. And what I call there is the increasing functionality. Functionality meaning the things we can do. And I think that's exactly what actual intelligence, what actual general intelligence is, is this ability to look at your environment, get the information, but that's not enough, then extract knowledge out of the information. What I mean by knowledge, if you really know, if you understand some information, it means you can apply it. So you can do some functionality. An easy example that when I talk about this I like giving is the idea of the anecdote we have with Newton and the apple falling from the tree.

Everybody sees the apple falling. That's not new information for anybody. We know that if you drop something, it's going to fall. You don't discover that. What Newton did is looked at this information and tried to get knowledge out of it.

Robert J. Marks:

I like that example of the apple falling.

Georgios Mappouras:

Yeah, I think it's easy to communicate it and that's why I use it. And what he said is that objects attract each other even if we know that today that's not really true because of Einstein. He did that. And then from that it's not just he did that observation, but he got some new knowledge. What does it mean? He could apply that knowledge and we can applied it to better describe how the planets move. So there's some functionality that from this understanding.

Robert J. Marks:

Well, let me first of all chime in. I think one of the definitions of artificial general intelligence is having all of the intelligence of all the libraries in the world available to you at the touch or the query of one of these large language models. The question is whether general intelligence generates what I would call creativity, Newton's observation, it's a myth, but it's a myth of him looking at the apple falling from the tree and then being able to be creative and actually extract from that I think is something that AGI probably will never do. But I think that that's a good test for AGI above and beyond just having access to all of the world's libraries. Now you mentioned the idea of functional information and non-functional information, and could you elaborate on that and the difference between the two? You kind of touched on it, but if you would drill a little bit deeper, that would be good.

Georgios Mappouras:

Right, right, right. So I think that's exactly, if we go back to this example, the information that thinks if I drop something, it's going to drop on the ground. If I hold something high, I leave it, it drops on the ground. That information exists. But what it actually means, not everybody knows. So this information is non-functional as in I cannot get some functionality out of it. I don't know what to do with this information. But then you can think it's like a chemical reaction where the true intelligence, an AGI, a general intelligence entity, can look at that information and as a byproduct it moves this non-functional information and transforms its to functional information. And the byproduct of that is new knowledge, new functionality.

So this is what you can think of AGI as a system that takes non-functional information, some information that I don't know what to do with it, and outputs functional information, meaning information that now I know what it means. I know that the reason things drop is because objects attract each other. And now out of this process, the byproduct is new knowledge, new functionality. In this case, it's the functionality of I'm able to predict how things move in space, for example.

Robert J. Marks:

So I guess I would say for the non-functional information that this requires creativity, that in order for Newton to extrapolate the idea of Newtonian laws of physics from an apple falling from a tree required a lot of creativity. Newton famously said that what he did was because he stood on the shoulders of giants and that standing on the shoulder of giants is all of the corpus of material in the world. And I think he had a library of something like 2,000 books, and those were the giants. He was probably referring to people like Galileo and people that preceded him. But the idea of creativity was not contained in those books, but it was actually the top of the mountain. He got to the top of the mountain and then he added onto it. He came up with this idea of creativity. And I think that that's what you mean by the non-functional information. I think that does require creativity. Would you agree?

Georgios Mappouras:

I agree, yeah. I think it's a good definition for creativity, even in terms that, creativity, even the term information, a lot of times they have some ambiguity inside of them. What do we mean by creativity? So that's what I like to do in this paper. I try to define those things, but rather than reusing terms like let's say creativity, AGI, new knowledge, all of them, in the end I try to find, let's say, description that fits all these things. And I think that's what I try to do with this non-functional information and functional meaning, the same information, the same data may be non-functional for somebody because it cannot interpret them, doesn't know what to do with them. But at the same time, for somebody else might be a functional information because they know how to use it. And the idea is that, for example, this comes back to the question, how do you detect it?

Let's say we have two humans, right? One has a hammer and knows how to use it and one has a screwdriver and knows how to use it and they can teach each other. So when they teach each other, what they do is that they exchange functional information. They say, "Look at this tool, this is how you use it." So they exchange functional information, but then, okay, now both of them know how to use both the hammer and the screwdriver. Is that it? So after all humans exchange all their ideas, is that it, is information done? Can we learn anything new? That's a system. For example, if that was a case, that would be a system that is not really intelligent, it's not really general intelligent, truly general intelligent system would be able to exactly what you said, reach the mountain and then add something to the top, look at the information, kind of extract even more knowledge, even more functionality.

Robert J. Marks:

I think that's good. I think it might be difficult. I'm a proponent of the Lovelace Test and the idea of whether it's new or not is sometimes very difficult because what if it comes up with an idea and this large language model or this transformer method, whatever you want to call it, has digested all of the pros in the world and maybe it comes up with an answer that is somewhere that it learned, but you don't know about it and you look at that and say, "Oh, that's creative, that's kind of new." I think it might be difficult to determine. I think that this is a problem of differentiating a non-functional information and creativity even in the Lovelace Test. What do you think about that?

Georgios Mappouras:

I agree. I agree. And that's why I call it the general intelligence threshold.

Robert J. Marks:

So define that. I mean that's the title of your paper. Define what the general intelligence threshold is.

Georgios Mappouras:

So exactly. So I don't think intelligence means really inventing something new per se. As you said, it might be out there somewhere, right? But if you can prove that for example, you have a system that maybe you taught it a specific, let's say class of functions. So you have a group of functions, a group of knowledge, and maybe you also give it enough information. So you tell it, for example, objects fall from the trees and then you expect it to come with a new knowledge, with a new functionality. So that's exactly why it is a threshold because this is actually an important part. Typically, when people try to measure intelligence, they have some, let's call it, they have some kind of levels and they say, "Okay, where are you in this bar from zero to hundred," let's say, which is very hard to define where is the threshold where for me, rather than trying to give you a score, I just have this threshold that is like, "Can you pass it or not?"

And this is the main difference. I don't try to give you a score. I'm trying to say, "Can you produce this thing? Can you showcase that you can do this?" That means functionality, that means, sorry, general intelligence. And so what I ask for a system is if I give you specific amount of knowledge, specific amount of functionality, and I give you enough information, what I call non-functional information. So information that you might not know how to use yet, but there's enough information there for you to extract new knowledge out of it. So this exactly the example, imagine Newton having all the library of physics behind him, so it has whatever he needs, all the information, all the experiments he has for is there and he has the current knowledge. Can he produce something? Let's say we have a system where we train it with what Newton knew at that time.

Can he produce the new knowledge, even if it's knowledge is known now, it doesn't matter if you can show that without looking, without looking at today's library, by just giving him the library of that time, can he produce new information? So this is the difference, doesn't have to be truly new. The system has to show that can produce new knowledge. An example would be, let's say, I assume we find an alien civilization, and maybe because typically people think about aliens, they're more progressed than us. Let's assume the opposite. They're less progress on us. They're still using, I don't know, tools like living in caves and whatever. How could we tell if they're intelligent, truly intelligent, general intelligent, we can see this progression. Can we see that there's new functionality before that they didn't have fire, now they use fire, right? Oh, now they also went from using stone to oil. If they can have this progression of new functionalities that shows general intelligence.

Robert J. Marks:

Okay, this is interesting. There was a movie, I forget the name of it though, where we were looking at extraterrestrial sort of life and the first thing that they did from a distant planet was kind of elucidate the prime numbers. And they talked about a number of the prime numbers. And all of a sudden we said, "Whatever is generating this must be intelligent." So that was the first step. Let me give you some examples of what I think would be good non-functional information. And I think this would be indisputable, and it kind of answers my question of whether or not this is hidden somewhere in the training data of the transformer. Recently, very recently, there was this 17-year-old homeschooled girl named Hannah Cairo, and she proved something called the Mizohata-Takeuchi conjecture. Actually she disproved it, but there are a lot of problems like this that are open in mathematics and you can be assured that nowhere has anybody ever solved this.

And then a few years ago, 2003, Grigori Perelman solved something called the Poincare conjecture. And this is I think a 40-year-old problem, not 40-year-old, I think the Mizohata-Takeuchi problem was 40-years-old, but the Poincare was hundreds a year old. And then Andrew Wiles in 1994 proved Fermat's Last Theorem. This is over 350 years old and nobody has ever checked it out. So there's all of these unproved open problems in math, including the Twin Prime Conjecture, the Collatz conjecture, the Riemann hypothesis, and any of these which they are solved by AI, would I think be indisputable proof that AI was creative. And of course what you would do is, like you said, you would give it all of the knowledge of mathematics. And if you go into ChatGPT, and I'm sure with the other large language models, they could do mathematics now, they can solve differential equations and Laplace transforms and do stochastic process analysis. It's really scary. So anyway, that's my proposal of something which is indisputable.

Georgios Mappouras:

Yeah, I agree. I agree with that view. It's always like, because they keep learning, there's always a question mark, did somebody else have this proof first if found it on the internet? And that's why it's as a test, you really want to test it. Sometimes we think if it's the model that is intelligent, the software, if it's actual the software, then you don't have to keep training. You can train it with less data and see if you can produce something we already know. The reason that makes it easier is because we know what knowledge can come from this information. For example, if I give you a book that only talks about, let's say, medicine and how to do a specific therapy, let's say, you should be able to get that knowledge out of that book.

So by saying, "Okay, if the model is actually intelligent, we don't need to always keep it at the edge of today's science. We can actually go backwards," say, "Okay, the science at some point was here. Can you produce what science produce after that?" I know what I expect. I know where you are right now. And disconnected, of course, disconnected from the internet. So this is actually one of my, when I proposed the Turing Test 2.0, this is one of the requirements is that you provide some knowledge to the system, but then there should be no other external functionality that comes from anywhere because that's how you can really tell, "Okay, nobody else gave this information to the system."

Robert J. Marks:

Great. You talk in your paper about Searle's Chinese Room, which I think is a smack down of the ability for computers to understand what they're doing. I think that the general idea, but you use it in kind of a different way. Could you explain the Searle's Chinese Room and how you use it in your argument about your theory?

Georgios Mappouras:

Yeah, actually this is a funny story because it kind of all came together. I was on YouTube listening to random videos, and I came across John Lennox and he was describing the argument and it was the first time I was hearing and I was like, "Wow, this is mind-boggling."

Robert J. Marks:

Yeah, John Lennox, for those who don't know, was a professor of mathematics at... I think it was Cambridge, wasn't it? Cambridge University. And he since retired, and right now he's kind of a Christian apologist, but he delves into artificial intelligence quite a bit. So anyway, continue.

Georgios Mappouras:

So for people that might not be familiar, the Chinese room argument is that they have a person that doesn't know how to speak Chinese. You close him into a room where he has no access to the real world, and you give him some specific instructions that explain him how to manipulate Chinese characters in order to form responses. And you have people outside that write Chinese characters, they slip a note under the door, the guy inside the room gets the note, he follows the algorithm, writes a response back without knowing what is actually writing slips the response back. And for the people outside the room, it looks like they can communicate. And John Lennox was talking about it and he was saying why this is not truly aware of what he's writing, doesn't know what he's writing. And then I heard the counter-argument. So again, listening to podcasts and stuff, I heard the counter-argument.

That's why I'm saying everything feels like a click together because it's like if maybe the YouTube algorithm, I don't know, was showing me what I needed to hear. And the counter-argument was like, "Wait a second. Yeah, maybe the person inside the room doesn't know, but the whole system, the person in the room plus the algorithmic instructions of how to manipulate the characters, plus input/output maybe that's what makes it understand the language." And that put me in thoughts of like, "Wait a second, is that true?" Kind of sounds true, but it doesn't satisfy me in terms of intellectually. So I was like, "Okay, what if in the instructions, so what if somebody slips in a node and in the node explains him how to escape them? Will he be able to act upon this information?" And that's where the idea came from. Can I actually extract new functionality out of that information?

The guy actually in the room might be talking to you and seems like he understands Chinese, but he cannot act in a way that knows knowledge of the subject. Because if he tells him, for example, "Oh, the password in order to unlock the store is this thing, just put it and get out to free yourself," it won't be able to act on it just because it doesn't really understand Chinese. So that's where the paper came from actually through this observation that it's one thing to have information, it's another thing for the information to have knowledge of what the information means. So information can be non-functional, I have it, I have the Chinese document in front of me, but it's non-functional because I cannot understand it. It becomes functional as long as I can understand it and get knowledge out of it. And now I know how to use it in this case by opening the door and the guy goes out of the room and is free. And this is what inspired me to make this observation.

Robert J. Marks:

Okay. Okay. Let me ask you, the final question, at least for this podcast is do you believe your Turing Test 2.0 has been passed? And what do you think of the chances of it being passed if your answer to the first question is no?

Georgios Mappouras:

Yeah, so I don't think part of the paper wasn't... I didn't, let's say, thoroughly tested. It was more like, "Here is a test." But as much as I tested it, it didn't seem like today's AI can pass it. And it's important caveat is that a system in order to pass, it doesn't have to constantly show that it can do that, right? It can just show once, like humans, for example, they can show once maybe in one aspect, maybe I'm good at math, but not good at physics.

Robert J. Marks:

Sorry to interrupt. Let me interrupt your flow and then I want you to go back to it. But I would call this a flash of genius. That's what Roger Penrose called it, a flash of genius. And it used to be in the United States in order to get a US patent that you had to have a flash of genius. They no longer have that requirement. But I think the flashes of genius are what you're talking about in terms of the non-functional information.

Georgios Mappouras:

Exactly. Yeah. I think that's a very nice term, and you don't have to always show it. You have to show it once, and that's enough to show that you are general intelligent. And again, all the tests I did and all that, as much as I could look at it does not look like they pass any type of test that is falls under the

Turing Test 2.0. and whatever pass it. I don't see a path to it, let's say, right? So currently I don't see a path how that happened. Yeah.

Robert J. Marks:

Good. Yeah, I guess I agree with you. Thank you, George. This was fun. We're going to continue this in a subsequent podcast, but we've been talking to Dr. Georgios Mappouras, who has written a fascinating paper called The Turing Test 2.0, the General Intelligence Threshold, and we'll continue this next time on Mind Matters News. Until then, be of good cheer.

Announcer:

This has been Mind Matters News with your host Robert J. Marks. Explore More at mindmatters.ai. That's mindmatters.ai. Mind Matters News is directed and edited by Austin Egbert. The opinions expressed on this program are solely those of the speakers.