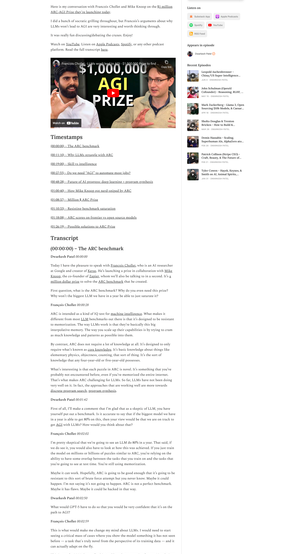
Francois Chollet, Mike Knoop - LLMs won't lead to AGI - $1,000,000 Prize to find true solution



Timestamps

00:00:00 - The ARC benchmark

00:13:00 - Why LLMs struggle with ARC

00:19:00 - Skill vs intelligence

00:27:10 - Do we need "System 2" to augment deep learning?

00:34:00 - Future of AI progress: deep learning + program synthesis

01:00:00 - How Mike Knoop got into AGI and founded ARC

01:06:21 - Millions of $ ARC Prize

01:10:21 - Recursive benchmark generation

01:18:00 - ARC success on frontier LLM context windows

01:26:10 - Possible solutions to ARC Prize

Transcript

(00:00:00) - The ARC benchmark

Dwarkesh Patel 00:00:00

Today I have the pleasure to speak with Francois Chollet, who is an AI researcher at Google and the creator of Keras. He's also the author of the influential paper, On Measure of Intelligence, and he's also the author of a recent, it's a million dollar prize to solve the ARC benchmark that he created.

First question, what is the ARC benchmark? Why do you even think it's a prize? Why won't the biggest LLM be here in a year be able to just solve it?

Francois Chollet 00:00:38

ARC is intended as a kind of IQ test for machine intelligence. What makes it different from most LLM benchmarks out there is that it's designed to be resistant to memorization. The way LLMs work is that they're basically this big interpolation memory. The way you scale up their capabilities is by trying to cram as much knowledge and patterns as possible into them.

By contrast, ARC does not require a lot of knowledge at all. It's designed to only require what's known as core knowledge. It's basic knowledge about things like elementary physics, elementary arithmetic, elementary geometry, and it's basic knowledge that you have near-old of five-year-old prepackaged.

What's interesting is that each puzzle in ARC is novel. It's something you've probably not encountered before, even if you've memorized the entire internet. That's what makes ARC challenging for LLMs. So far, LLMs have not been doing very well on it, but, the approaches that are working well are more towards discrete program search, neural synthesis.

Dwarkesh Patel 00:01:47

First of all, I'll make a comment that I'm glad that as a skeptic of LLM, you have yourself put out a benchmark. It is accurate to say that if the biggest model that he is just able to get 80% on this, that's a year time model be that we are on track to get AGI with LLMs? How would you think about that?

Francois Chollet 00:02:02

I'm pretty skeptical that we're going to see an LLM do 80% in a year. That said, if we do see it, you would also have to look at how this was achieved. If you just cram as many examples into your LLM as you can, hoping that some of them will be able to fake some reasoning between the tasks that you cram in and the tasks that you're going to see at test time. That is still using memorization.

Maybe it can work. Hopefully, ARC is going to be pushed enough that it's going to be resistant to this sort of brute force strategy but you never know. Maybe it could happen. I'm not saying it's not going to happen. ARC is not a perfect benchmark. Maybe it has flaws. Maybe it could be hacked in that way.

Dwarkesh Patel 00:02:50

What would GPT-5 have to do or what do you would be very confident that it's on the path to AGI?

Francois Chollet 00:02:59

This is what would make me change my mind about LLMs. I would need to see a model that is able to learn new skills and to pick up new skills on the fly while only being trained on new data. That would be the proper way of training data. I and it can actually adapt on the fly.

This is not how LLMs have really this would catch my attention for any AI technique out there. If I can see the ability to adapt to novelty on the fly and pick up new skills, that would be extremely interested. I would think this is on the path to AGI.

Dwarkesh Patel 00:03:39

The advantage they have is that they do get to see everything. Maybe I'll take issue with how much they are relying on that, but obviously they're relying on that more than humans do. They do have so much in distribution, to the extent that we have trouble distinguishing whether an example is in distribution or not.

If they have everything in distribution, then they can do everything that we can do because we can only do what is in distribution. Why do you think that being in distribution for them? Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:04:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! We can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:04:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:05:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Dwarkesh Patel 00:06:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:06:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:07:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:08:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:08:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:09:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:10:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:10:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:11:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:12:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:12:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:13:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:14:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:14:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:15:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:16:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:16:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:17:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why we, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:18:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:18:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:19:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:20:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:20:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:21:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:22:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:22:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:23:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:24:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:24:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:25:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:26:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:26:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:27:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:28:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:28:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:29:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:30:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:30:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:31:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:32:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:32:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:33:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:34:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:34:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:35:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:36:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:36:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:37:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:38:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:38:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:39:32

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:40:12

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have developed intelligence in the first place. If the world was static, there would be no need for intelligence. We could just rely on distribution for dumb! Why can't we just leverage the fact that they do get to see everything?

Francois Chollet 00:40:52

Basically you're asking what's the difference between actual intelligence ... the ability to adapt to things you've not been prepared for ... and pure memorization, like reciting what you've seen before.

The crux of intelligence is novelty. The big difference is that you can never pick up on everything that you might see or even have the novelty adapted to all the time. It's just the fact that the space of possible tasks is infinite. If you've trained an millions of them, you've only seen zero percent of the total space. It's also the fact that the world is changing every day.

This is why, the human species, have

example. They would give the model, in context, the grammar book and the dictionary of a language that has fewer than 200 living speakers. It's not in the pre-training data. You just give it the dictionary and it basically is able to speak this language and translate to it, including the complex and organic ways in which languages are structured.

If you showed me a dictionary from English to Spanish, I'm not going to be able to pick up how to structure sentences and how to say things in Spanish. Because of the representations that it has gained through this pre-training, it is able to now learn a new language extremely efficiently. Doesn't that show that this kind of pre-training actually does increase your ability to learn new tasks?

François Chollet *00:25:58*

If you were right, LLMs would do really well on ARC puzzles because ARC puzzles are not complex. Each one of them requires very little knowledge. Each one of them is very low on complexity. You don't need to think very hard about it. They're actually extremely obvious for human

Even children can do them but LLMs cannot. Even LLMs that have 100,000x more knowledge than you do still cannot. The only thing that makes ARC special is that it was designed with this intent to resist memorization. This is the only thing. This is the huge blocker for LLM performance.

If you look at LLMs closely, it's pretty obvious that they're not really synthesizing new programs on the fly to solve the task that they're faced with. They're very much reapplying things that they've stored in memory. For instance, one thing that's very striking is that LLMs can solve a Caesar cipher, transposing letters to code a message. That's a very complex algorithm, but it comes up quite a bit on the internet. They've basically memorized it.

What's really interesting is that they can do it for a transposition length of like three or five, because those are very common numbers in examples provided on the internet. If you try to do it with an arbitrary number like nine, it's going to fail. It does not encode the generalized form of the algorithm, but only specific cases. It has memorized specific cases of the algorithm. If it could actually synthesize on the fly the solver algorithm, then the value of *n* would not matter at all, because it does not increase the problem complexity.

Dwarkesh Patel *00:26:48*

I think this is true of humans as well.

François Chollet *00:26:51*

Humans use memorization pattern matching all the time, of course, but humans are not limited to memorization pattern matching. They have this very unique ability to adapt to new situations on the fly. This is exactly what enables you to navigate every new day in your life.

Dwarkesh Patel *00:27:07*

There was some study that chess grandmasters will perform very well within the context of the moves that—

François Chollet *00:27:14*

That's an excellent example because chess, at the highest level, is all about memorization, chess memorization.

Dwarkesh Patel *00:27:19*

What is your explanation for the original question of why Gemini 1.5 was able, in context, to learn a language, including the complex grammar structure? Doesn't that show that they can pick up new knowledge?

François Chollet *00:27:35*

I would assume that it has simply mined from its extremely extensive, unimaginably vast training data. It has mined the required template and then it's just reusing it. We know that LLMs have a very poor ability to synthesize new program templates like this on the fly or even adapt existing ones. They're very much limited to fetching.

(00:27:55) - Do we need "AGI" to automate most jobs?

Dwarkesh Patel *00:27:55*

Suppose there's a programmer at Google. They go into the office in the morning. At what point are they doing something that 100% cannot be due to fetching some template?

Suppose they were an LLM. What could they not do if they had only fetched some template from their program? At what point do they have to use this so-called extreme generalization capability?

François Chollet *00:28:12*

Forget about Google software developers. For every human, every day of their lives is full of novel things that they've not been prepared for. You cannot navigate your life based on memorization alone. It's impossible.

Dwarkesh Patel *00:28:25*

It seems like you also agree they're not doing just "memorization." It seems like you're saying they're less capable of generalization. I'm just curious about the kind of generalization they do.

If you get into the office and you try to do this kind of generalization, you're going to fail at your job. Let's say you're a programmer. What is the first point when you try to do that kind of generalization, you would lose your job because you can't do the extreme generalization?

François Chollet *00:28:51*

Take this situation, for instance. You've never been here in this room. Maybe you've been in this city a few times. There's a fair amount of novelty. You've never been interviewing me. There's a fair amount of novelty in every hour of every day in your life. By and large, it's in fact more novelty than any LLM could handle. If you just put an LLM in a robot, it could not be doing all the things that you've been doing today.

Take self-driving cars, for instance. You take a self-driving car operating in the Bay Area. Do you think you could just drop it in New York City or drop it in London, where people drive on the left? No, it's going to fail. Not only can it not generalize to a change in driving rules, but you cannot even make it generalize to a new city. It needs to be trained on each specific environment.

Dwarkesh Patel *00:29:55*

I agree that self-driving cars aren't AGI.

François Chollet *00:29:58*

But it's the same type of model. They're transformers as well. It's the same architecture.

Dwarkesh Patel *00:30:02*

I don't know. Apes also have brains with neurons in them, but they're less intelligent because they're smaller.

We can get into that. I still don't understand this concrete thing. We also need training. That's why education exists. That's why we had to spend the first 18 years of our life doing drills.

François Chollet *00:30:19*

We have a memory, but we are not a memory. We are not limited to just a memory.

Dwarkesh Patel *00:30:24*

I'm denying the premise that that's the only thing these models are necessarily doing.

Suppose you just subbed out a remote work with an LLM and they're a programmer. What is the first point at which you realize this is not a human, this is an LLM?

François Chollet *00:30:40*

How about I just send them an ARC puzzle and see how they do?

Dwarkesh Patel *00:30:44*

No, like part of their job.

François Chollet *00:30:46*

You have to deal with novelty all the time.

Dwarkesh Patel *00:30:49*

Is there a world in which all the programmers are replaced and we're still saying, "ah, but they're only doing memorization-laden programming tasks." In that world, are they still producing a trillion dollars worth of output in the form of code?

François Chollet *00:31:03*

Software development is actually a pretty good example of a job where you're dealing with novelty all the time. If you're not, I'm not sure what you're doing.

I personally use generative AI very little in my software development job. Before LLMs, I was also using Stack Overflow very little. Some people maybe are just copy-pasting stuff from Stack Overflow, or nowadays copy-pasting stuff from an LLM.

Personally, I try to focus on problem-solving. The syntax is just a technical detail. What's really important is problem-solving. The essence of programming is engineering mental models and mental representations of the problem you're trying to solve.

Dwarkesh Patel *00:31:46*

We have many people who can interact with these systems themselves. You can go to ChatGPT and say, "here's a specification of the kind of program I want." They'll build it for you.

François Chollet *00:31:56*

As long as there are many examples of this program on GitHub, Stack Overflow, and so on, sure they will fetch the program for you from their memory.

Dwarkesh Patel *00:32:03*

But you can change arbitrary details. You can say, "I need it to work on this different kind of server."

François Chollet *00:32:09*

If that were true, there would be no software engineers today.

Dwarkesh Patel *00:32:12*

I agree we're not at a full AGI yet. These models have fewer than a trillion parameters. A human brain has somewhere on the order of 10-30 trillion synapses. If you were just doing some naive math, you're at least 10x underparameterized. I agree we're not there yet, but I'm confused about why we're not on the spectrum.

Yes, I agree that there are many kinds of generalization they can't do. But it seems like they're on this kind of smooth spectrum that we see even within humans. Some humans would have a hard time doing an ARC-type test. We see that based on the performance on Raven's progressive matrices-type IQ tests.

François Chollet *00:32:48*

I'm not a fan of IQ tests because, for the most part, you can train on IQ tests and get better at them. They're very much memorization-based. This is actually the main pitfall that ARC tries not to fall for.

Dwarkesh Patel *00:33:02*

Let's say all remote jobs are automated in the next five years. I mean at least the remote jobs that don't require you to be a sort of a service, like a salesperson, where you want the human to be talking. I mean more like programming.

In that world, would you say that that's not possible because a programmer needs to do many things that definitely require things that would not be in any pre-training corpus?

François Chollet *00:33:25*

Sure. In five years, there will be more software engineers than there are today, not fewer.

Dwarkesh Patel *00:33:30*

I'm still not sure. I studied computer science. If I had become a code monkey out of college, what would I be doing? I go to my job. My boss tells me to do something? When does he realize I'm an LLM, if I were an LLM?

François Chollet *00:33:46*

Probably on the first day. Again, if it were true that LLMs could generalize to novel problems like this — actually develop software to solve a problem they've never seen before — you would not need software engineers anymore.

If I look at how people are using LLMs in their software engineering job today, they're using it as a Stack Overflow replacement. They're using it as a way to copy-paste code snippets to perform very common actions. What they actually need is a database of code snippets. They don't actually need any of the abilities that actually make them software engineers.

Dwarkesh Patel *00:34:27*

Let's step back on interpolation. Why isn't creativity just interpolation in a higher dimension where — if we're going to use the ML language — a bigger model can learn a more complex manifold?

If you read a biography of a scientist, they're not zero-shotting new scientific theories. They're playing with existing ideas. They're trying to juxtapose them in their head. In the tree of intellectual descendants, they try out some slightly different evolutionary path. You sort of run the experiment there in terms of publishing the paper or whatever.

It seems like a similar kind of thing to what humans are doing. There's a higher level of generalization. Bigger and bigger models seem to be approaching higher and higher levels of generalization. GPT-2 couldn't do grade school-level math problems that required more generalization than it had the capability to do. GPT-3 and GPT-4 can.

François Chollet *00:35:32*

Not quite. GPT-4 has a higher degree of skill and a higher range of skills. It has the same degree of generalization.

Dwarkesh Patel *00:35:39*

I don't want to get into semantics here. Why can't creativity just be interpolation on a higher dimension?

François Chollet *00:35:49*

Interpolation can absolutely be creative. To your point, I do think that on some level humans also do a lot of memorization, reciting, pattern matching, and interpolation as well. It's very much a spectrum between pattern matching and true reasoning. Humans are never really at one end of the spectrum. They're never really doing pure pattern matching or pure reasoning. They're usually doing some mixture of both.

This is true even if you're doing something that seems very reasoning-heavy, like proving a mathematical theorem. As you're doing it, you're doing quite a bit of discrete search in your mind and quite a bit of actual reasoning. You're also very much guided by intuition and pattern matching. You're guided by the shape of proofs that you've seen before, by your knowledge of mathematics.

All of our thoughts, everything we do, is a mixture of interpolated memorization-based thinking, Type 1 thinking, and Type 2 thinking.

Dwarkesh Patel *00:36:55*

Why are bigger models more sample efficient?

François Chollet *00:36:59*

Because they have more reusable building blocks that they can lean on to pick up new patterns in their training data.

Dwarkesh Patel *00:37:09*

Does that pattern keep continuing as you keep getting bigger and bigger?

François Chollet *00:37:12*

It does to the extent that the new patterns you're giving the model to learn are a good match for what it has learned before. If you present something that's actually novel that is not in a steady distribution, like an ARC puzzle for instance, it will fail.

Dwarkesh Patel *00:37:25*

Let me make this claim. The program synthesis is a very useful intuition pump. Why can't this be the case for what's happening in the transformer?

The early layers are figuring out how to represent the inputting tokens. The middle layers do this kind of program search, program synthesis, and they combine the inputs to all the circuits in the model. They go from the low-level representation to a higher-level representation near the middle of the model. They use these programs. They combine these concepts. What comes out the other end is the reasoning based on that high-level intelligence.

François Chollet *00:38:01*

Possibly. Why not? But if these models were actually capable of synthesizing novel programs, however simple, they should be able to do ARC. Because for any ARC task, if you write down the solution program in Python, it's not a complex program. It's extremely simple. Humans can figure it out. Why can't LLMs do it?

Dwarkesh Patel *00:38:26*

That's a fair point. To turn the question around to you, suppose it's the case that in a year a multimodal model can solve ARC. Let's say it gets 80% or whatever the average human would get. Are we then on track for AGI?

François Chollet *00:38:43*

Quite possibly, yes. Honestly, what I would like to see is an LLM-type model solving ARC at 80%, but after having only been trained on core knowledge-related stuff.

Dwarkesh Patel *00:38:59*

But human kids, we're necessarily just trained on what we have in our genes…

François Chollet *00:39:04*

Let me rephrase that. I want it to be only trained on information that is not explicitly trying to anticipate what's going to be in the ARC test set.

Dwarkesh Patel *00:39:15*

Isn't the whole point of ARC that you can't? It's a new type of intelligence test every single time?

François Chollet *00:39:21*

Yes, that is the point. If ARC were a perfect, flawless benchmark, it would be impossible to anticipate what's in the test set. ARC was released more than four years ago and so far it's been resistant to memorization. It has, to some extent, passed the test of time. But it's not perfect.

Let's say you try to make by hand hundreds of thousands of ARC tasks. You try to multiply them by programmatically generating variations. You end up with maybe hundreds of millions of tasks. Just by brute forcing the task space, there will be enough overlap between what you're trained on and what's in the test set that you can actually score very highly. With enough scale, you can always cheat.

Dwarkesh Patel *00:40:05*

If you can do this for every single thing that supposedly requires intelligence, then what good is intelligence? Apparently, you can just brute force intelligence.

François Chollet *00:40:12*

If the world, if your life, were a static distribution then sure, you could just brute force the space of possible behaviors. There are several metaphors for intelligence I like to use. One is that you can think of intelligence as a pathfinding algorithm in future situation space.

I don't know if you're familiar with RTS game development. You have a map, a 2D map, and you have partial information about it. There is some fog of war on your map. There are areas that you haven't explored yet. You know nothing about them. There are also areas that you've explored but you only know what they were like in the past. You don't know how they are like today.

Now, instead of thinking about a 2D map, think about the space of possible future situations that you might encounter and how they're connected to each other. Intelligence is a pathfinding algorithm. Once you set a goal, it will tell you how to get there optimally. Of course, it's constrained by the information you have. It cannot pathfind in an area that you know nothing about. It also cannot anticipate changes.

If you had complete information about the map, then you could solve the pathfinding problem by simply memorizing every possible path, every mapping from point A to point B. You could solve the problem with pure memory. The reason you cannot do that in real life is because you don't actually know what's going to happen in the future. Life is ever changing.

Dwarkesh Patel *00:41:56*

I feel like you're using words like "memorization," which we would never use for human children. If your kid learns to do algebra and then learns to do calculus, you wouldn't say they've memorized calculus. If they can solve any arbitrary algebraic problem, you wouldn't say they've memorized algebra. You'd say they've learned algebra.

François Chollet *00:42:11*

Humans are never really doing pure memorization or pure reasoning.

Dwarkesh Patel *00:42:15*

That's only because you're semantically labeling what the human does as skill. But it's a memorization when the exact same skill is done by the LLM, as you can measure by these benchmarks. You can just plug in any sort of math problem.

François Chollet *00:42:22*

Sometimes humans are doing the exact same as the LLM is doing. For instance, if you learn to add numbers you're memorizing an algorithm. You're memorizing a program and then you can reapply it. You are not synthesizing on the fly the addition program.

Dwarkesh Patel *00:42:38*

Obviously at some point, some human had to figure out how to do addition. A kid doesn't figure it out by starting from the axioms of set theory and going to how to do addition.

François Chollet *00:42:47*

What you learn in school is mostly memorization.

Dwarkesh Patel *00:42:50*

My claim is that these models are vastly underparameterized relative to how many flops, how many parameters, you have in the human brain. So it makes sense that they're not going to be coming up with new theorems like the smartest humans can. Most humans can't do that either. What most humans do sounds like something similar to what you are calling memorization, which is memorizing skills or memorizing techniques that you've learned. So it sounds like it's compatible.

Tell me if this is wrong. Is it compatible in your world if all the remote workers are gone but they're doing skills which we can potentially make synthetic data out of? We record every single remote worker's screen. We sort of understand the skills they're performing there. Now we've trained a model that can do all this. All the remote workers are unemployed. We're generating trillions of dollars of economic activity from AI remote workers. In that world, are we still in the memorization regime?

François Chollet *00:43:45*

Sure, with memorization you can automate almost anything as long as it's a static distribution, as long as you don't have to deal with change.

Dwarkesh Patel *00:43:54*

Are most jobs part of such a static distribution?

François Chollet *00:43:57*

Potentially, there are lots of things that you can automate. LLMs are an excellent tool for automation. But you have to understand that automation is not the same as intelligence. I'm not saying that LLMs are useless. I've been a huge proponent of deep learning for many years.

For many years, I've been saying two things. I've been saying that if you keep scaling up deep learning, it will keep paying off. At the same time I've been saying if you keep scaling up deep learning, this will not lead to AGI.

We can automate more and more things. Yes, this is economically valuable. Yes, potentially there are many jobs you could automate away like this. That would be economically valuable. You're still not going to have intelligence.

So you can ask, what does it matter if we can generate all this economic value? Maybe we don't need intelligence after all. You need intelligence the moment you have to deal with change, novelty, and uncertainty.

As long as you're in a space that can be exactly described in advance, you can just rely on pure memorization. In fact, you can always solve any problem. You can always display arbitrary levels of skills on any task without leveraging any intelligence whatsoever, as long as it is possible to describe the problem and its solution very, very precisely.

Dwarkesh Patel *00:45:17*

When they do deal with novelty, then you just call it interpolation.

François Chollet *00:45:21*

No, interpolation is not enough to deal with all kinds of novelty. If it were, then LLMs would be AGI.

Dwarkesh Patel *00:45:30*

I agree they're not AGI. I'm just trying to figure out if we're on the path to AGI. The crux here is that it seems to me that these things are on a spectrum and we're clearly covering the earliest part of the spectrum with LLMs.

François Chollet *00:45:43*

I think so.

Dwarkesh Patel *00:45:44*

Okay, interesting. Here's another thing that I think is evidence for this: grokking.

Clearly, even within deep learning, there's a difference between the memorization regime and the generalization regime. At first they'll just memorize the data set. If you're doing modular addition it's how to add digits. At some point, if you keep training on that, they'll learn the skill.

The fact that there is that distinction suggests that for the generalized circuit that deep learning can learn, there is a regime where it generalizes if you have an overparameterized model. We don't have that in comparison to all the tasks we want these models to do right now.

François Chollet *00:46:20*

Grokking is a very, very old phenomenon. We've been observing it for decades. It's basically an instance of the minimum description length principle. Given a problem, you can just memorize a pointwise input-to-output mapping, which is completely overfit.

It does not generalize at all, but it solves the problem on the trained data. From there, you can actually keep pouring it and adding more stuff, mapping simple and more compressed. At some point, it will generalize.

That's something called the minimum description length principle. It's like saying that the program that will generalize best is the shortest. It doesn't mean that you're doing anything other than memorization. You're doing memorization plus regularization.

Swarbrick Point 00:47:23

A.k.a. generalization?

François Chollet 00:47:25

Yeah, that absolutely leads to generalization.

Swarbrick Point 00:47:28

So you do that within-task stuff. The pattern you see here of meta-learning is that it's more efficient to train a program that can perform many skills rather than one skill. This is what this weight of Kaiser mentions. So if you get bigger and bigger in models, you would expect it to go up the hierarchy of generalization. It generalizes to a skill, then it generalizes across multiple skills.

François Chollet 00:47:48

That's correct. LLMs are not inherently large. They have only a fixed number of parameters. They have to compress their knowledge as much as possible. In practice, they are very large because they are trained on very large datasets. Because they have this need for compression, every time they're learning a new program they're going to try to express it in terms of existing bits and pieces of programs that they've already learned before.

Swarbrick Point 00:48:08

Isn't this general intuition?

François Chollet 00:48:10

Absolutely. Clearly LLMs have some degree of generalization. This is precisely why. It's because they have compression.

Swarbrick Point 00:48:19

Why is that intuitionally limited? At some point it has to hitch a higher level of generalization and a higher level, and then the highest level is the third intelligence.

(00:48:28) - Future of AI progress: deep learning + program synthesis

François Chollet 00:48:28

It's intuitionally limited because the substrate of your model is a big parametric curve. All you can do with this is local generalization. If you want to go beyond this towards broader or even extreme generalization, you have to move to a different type of model. My paradigm of choice is discrete program search, program synthesis.

If you want to understand that, you can sort of compare and contrast a little deep learning and discrete program search. Deep learning is parametric curve fitting. To express complex, smooth, non-trivial or discrete program structures. You've got a set of logical operators, like a domain-specific language. You're picking instead of it. You're reassembling that into a graph that's a program. That's actually very similar to a program you might write in Python or C++ and so on. We are doing machine learning, but we are doing it in a completely different way.

Deep learning is learning by fitting a parametric curve. Gradient descent is very computationally efficient because it's differentiable. It's differentiable because it's local, meaning that at any point, you can get the relation very quickly. But it's very data inefficient. In order to make it work, you need a dense sampling of the information space you're trying to learn about. It's also very brittle. It's very tightly limited to only generalizing within that data distribution. It's very brittle because this limitation is because your model is curve.

Meanwhile, if you look at discrete program search, the learning process is combinatorial search. You're just trying a bunch of programs until you find one that actually means your data. This process is extremely data efficient. You can learn a generalizable program from just one example, two examples. This is why it's very robust. It's, by the way, it's very computationally inefficient because you're running into combinatorial explosion, of course.

You can sort of see how deep learning and discrete program search have very complementary strengths, and limitations as well. Every limitation of deep learning has a corresponding strength in program synthesis and inversely. The path forward is going to be to merge the two.

Here's another way you can think about it. These parametric curves trained with gradient descent are great for fitting smooth, discrete, continuous, high-dimensional interpolation, intuition, memorization, etc. Discrete program search is a great fit for reasoning, for abstraction, for anything that requires, for example, a recursive model that machine can run on two examples, like for an ARC puzzle for instance.

Humans are never doing pure System 1 or pure System 2. They're always mixing and matching both. Right now, we have all the tools for System 1. We have nothing for System 2. The way forward is to create a hybrid system.

The form it is going to take is mostly System 2. The main structure is going to be a discrete program search system. You're going to fit the fundamental limitation of discrete program search, which is combinatorial explosion, with deep learning. You're going to leverage deep learning to guide and to provide intuition to program search, to prune the program search.

There's very similar to what you see when you're playing chess or when you're trying to solve a difficult math problem. You have your intuition, your gut feeling, and you with some intuition about the shape of the solution. That's very much something you can get via a deep learning model. Deep learning models are very much like intuition machines. They're pattern matching machines.

You zoom from this shape of the solution, and then you're going to do actual reasoning, actual discrete search. Deep learning is going to be, in this case, they're not going to try doing reasoning. They're not going to try and assume that learning model for suggestions. It'll be like, There's the most likely next step. Here's where in the graph you should be going. You can also ask another deep learning model for feedback like "Isn't, here's what I have so far. Is it looking good? Is it not looking good? Should you continue?" And so on. Deep learning is going to be the key but you want to make it dramatically better, robust, and much more efficient, by leveraging deep learning.

By the way, another thing that you can say deep learning for is of course things like common sense knowledge and knowledge in general. You're going to end up with this sort of system where you have this on the synthesis engine that can adapt to new situations.

The way it adapts is that it's going to fetch from a bank of patterns, modules that could be themselves curves, differentiable modules, and some others that could be discrete. It's going to assemble, reassemble, on the fly, a new model, a new program. For every new situation you might be faced with, it's going to give you a generalizable model that was synthesized using very, very little data. Something like this would solve ARC.

François Chollet 00:54:18

That's actually a really interesting prompt. There's an interesting core here. I talk to my friends who are extremely optimistic about LLMs and expect AGI within the next couple of years. In some sense, they also agree that scaling is not all you need but that the rest of the program is underspecified and enabled by scaling. You still need to add the System 2 and the core intuition on top of that models.

Their perspective is that it's relatively straightforward to do that because you have the LLM, you have the intuition engine, and it's very easy to plug something on top of it to just slowing through methods. You need some more deliberate way in which it engages with the material it learns. In context learning is extremely simple efficient. You would, I think for the foundation, you would the model in order to through the things it sees and then add it back to the weights.

As far as the System 2 part, they talk about adding some kind of RL wrap so that it is encouraged to proceed on the reasoning traces that end up being correct. They think that's a relatively straightforward stuff that will be added within the next couple of years.

François Chollet 00:55:32

That's an empirical question on we'll see.

Swarbrick Point 00:55:35

I assume your intuition is not that. I'm curious why.

François Chollet 00:55:37

My intuition is that this whole System 2 architecture is the hard part. It's the very hard, hard intellectual path. Scaling up the transformers seemed to the easy part. It's literally just a big curve. All you need is more data. It's an interpolative representation of a data set. That's the easy part.

The hard part is the architecture of intelligence. Memory and intelligence are separate components. We have the memory. We don't have the intelligence yet. I agree with you that having the memory is actually very useful. If you had the intelligence, you would be able to plug it to an enormous memory. I would say that useful because it could not learn enough material to work from.

Swarbrick Point 00:56:21

Former guest Dwarkesh Patel proposed an alternative hypothesis that intelligence is just a function of memory and that if you had a very, very large memory, you could access sufficiently complex efficiency. You can just look at a few cases and figure out what was the moderate. But who is also because he has learned higher-level abstractions. It's attractive to some fundamental sense.

Here's one way to ask the question. In the brain, supposedly no dry program synthesis, but is it just compute connected to each other. Physically, it's go to be that you just query the right circuit, right?

François Chollet 00:57:00

You are, yeah. It's a matter of degree.

Swarbrick Point 00:57:08

Training in the environment that human ancestors were trained in means you have those circuits. If you train on the same kinds of outputs that humans produce — the same environment, the same kind of circuits — wouldn't that train the same thing that is inherent human brain?

François Chollet 00:57:19

It's a matter of degree. If you have a system that has a memory and is only capable of local generalization, then, yes, you can, in principle, learn, for instance, the ARC training data and the memory plus the ability to search to gain some depth to achieve broader and even extreme generalization.

One of my favorite psychologists is Isaac Asimov, the founder of developmental psychology. He had a very good point about intelligence. He said, "Intelligence is what you use when you don't know what to do." As a human living your life, in most situations you already know what to do because you've been in this situation before. You already have the answer.

You're only going to need to use intelligence when you're faced with novelty, with something you've never encountered before. In that case, you're going to be able to by your own life experience or your evolutionary history. This day that you're living right now is different in some important ways from every day you've lived before. It's also different from any day ever lived by any of your ancestors. You're still capable of being functional. How is that possible?

Swarbrick Point 00:58:09

I'm not denying that generalization is extremely important and the basis for intelligence. That's not the crux. The crux is how much of what is happening in the models.

Okay, let me ask a separate question about the difference in intelligence between humans. Maybe because of the reasons you mentioned, the intelligence now, we have the memory. And that's clearly there's a difference in intelligence between different humans.

What is your explanation for what's going on there? That's a sort of complicated with my mind. There's a spectrum of generality and these models are climbing up on a human level. Even some humans haven't even climbed up to the Einstein level or the Newton level.

François Chollet 00:59:04

That's a great question. There's extensive evidence for differences in intelligence are mostly genetic in origin. This means that if you take someone who is very low intelligence, there is an amount of training data you can expose that person to that would make their become Einstein. The genes in the fact that you really need a certain amount. You need a latent algorithm. More training data is not in fact all you need.

Swarbrick Point 01:00:02

I think I agree with that. I might phrase it in this way. The people who are smarter have better, more efficient, more powerful, more, whatever, more, better, better, is more efficient. Maybe they have greater density of firing.

Some part of the story is scaling. There's some correlation between brain size and intelligence. Within the context of "scaling", LLMs, people talk about architectural improvements. A model like Gemini 1.5 Flash performs as well as GPT-4 did when GPT-4 was released a year ago. So is it I know cheaper to recapture that of the scaling story is that we're in like extremely low hanging fruit territory when it comes to these architectural improvements.

(01:00:40) - How Mike Knoop got nerd-sniped by ARC

Swarbrick Point 01:00:40

We're back now with the co-founder of Zapier, Mike Knoop. You're leading this prize and you're running this prize with Francois. Tell me about how this came together. What prompted you guys to launch this prize?

Mike Knoop 01:00:52

I've been AI curious for 10 years. I co-founded Zapier and I've been running it for the last 10 years.

I've been interested in your work during COVID. I went down the rabbit hole. I had a lot of free time. It was right after you'd published your paper, On the Measure of Intelligence. You introduced the concept of ACI and that this efficiency of skill acquisition is the right definition, and the ARC puzzles.

I don't think the first Kaggle contest had been done yet. It was still starting. It was interesting but I just pushed the idea I had bigger feel to it to Zapier. We were in the middle of the big narrowing of trying to get us our second product.

It was January 2020 when the rabbit of Kaggle contest came out. That really sniped me in the program. I was given a little presentation to Zapier on the Kaggle contest. I'm a competitive, I'm a competitive person. I'm a competitive, I'm a competitive, shocking to me in terms of all those latent capabilities that I, like, have that I didn't expect they had.

I actually gave up my time team only. I was running half the company at that point. I was back to being an individual contributor and just doing AI research. I started a Slack. We started a Slack. Eventually, that led to Zapier actually, I'm looking back, I guess. I had expected to use this innovation effect for AGI and COVID-19.

When I looked at the teams and the progress over the last two years, I was very, I was shocked at how small the scale was. Little tiny effective progress towards it. It felt like it really important and, As I speak the last year spinning people about it to my network and community, very few people even know it existed. If it's right that this is the, I'm convinced this is the path to AGI, I'm very surprised that there's not more attention and more people who are currently obsessed. Or AGI — they don't even know about this thing.

I had my own idea on how to fix it. I wanted to work. I'm working nights and weekends on that. I flew up to meet Francois earlier this year to spin him and show him my ideas. Ultimately I realized that why more people didn't know about ARC? You should actually answer that. It's a really interesting question. Why don't you think more people know about ARC?

François Chollet 01:03:17

Benchmarks that gain traction in the research community are benchmarks that are already fairly crackable. The domain is that some research group is going to make a breakthrough and they're going to, they're going to, they're going to, they're going to, they're going to get follow-up papers with people trying to beat the first team and so on.

The thing that was really happened for ARC because ARC is actually very hard for existing AI technologies. ARC requires you to see very new ideas. That's very much the point. The point is to ensure that you should just be able to apply existing technology and solve ARC. The point is that existing technology has reached a plateau. If you want to make progress, you need to come up with new ideas. And these ideas are unprecedented or even before, you need to try new ideas.

ARC is not just meant to be this sort of measure of how close we are to ACI. It's also meant to be a source of inspiration. I want researchers to look at those puzzles and be like, "Hey, it's really strange that these puzzles are so simple and most humans can just do them very quickly. Why is it so hard for existing systems? Why is it so hard for LLMs and so on?"

This is now for LLMs, but ARC was actually released before LLMs were really a thing. The only thing that came to special at the time was GPT-2 was designed to be a natural language processing. It's a natural language benchmark. Yeah, and Cassie's in general, shows that is actually resistant to memorization.

Mike Knoop 01:04:34

This is what said-sniped me. I went and took a bunch of the puzzles myself. I'm, I'm, I'm,

end of that spectrum. You scale up the memorization by a massive factor and you're doing much better, which tells you that the same thing, just different scale of the spectrum.

I think where you're going to get the most value for your compute cycles is somewhere in between. You want to leverage memorization to build up a richer, more useful bank of primitive programs. You don't want them to be hard-coded like what we saw for the typical RTS. You want them to be learned from examples. You also want to do some degree of deep search. As long as you're only doing very shallow search, you are limited to local generalization. If you want to generalize further and more broadly, depth of search is going to be critical.

Dwarkesh Patel *01:16:26*

I might argue that the reason that he had to rely so heavily on the synthetic data was because he used a 3M million parameter model. The Kaggle competition at the time required him to use a P100-GPU which has like a tenth or something of the flops of an H100.

For context for the listeners, the frontier labs would rely on literally a thousand times bigger than this. For your competition, submissions can't make any API calls, can't go online, and have to run on NVIDIA Tesla P100. It's significantly less powerful.

Mike Knoop *01:17:20*

There's basically a 12 hour runtime limit. There's a forcing function of efficiency in the eval.

François Chollet *01:17:24*

But here's the thing, you only have 100 test tasks. The amount of computing available for each task is actually quite a bit, especially if you contrast that with the simplicity of each task.

Dwarkesh Patel *01:17:35*

Basically, it would be 7 minutes per task. People who have tried to do these estimates of how many flops does a human brain have. You can take them with a grain of salt but as a sort of anchor, it's basically the amount of flops an H100 has.

Maybe you would argue that a human brain can solve this question in faster than 7 minutes. Maybe you need a tenth of the compute, you should be able to do it in seven minutes. Obviously we have been their predecessor of their access memory in the brain and these 29 GB or whatever in the H100.

(01:18:08) - ARC scores on frontier vs open source models

Dwarkesh Patel *01:18:08*

The broader point is that I wish there were a way to also see this price with some sort of scaffolding on the biggest models, as a way to see whether scaling is the path to solving ARC.

François Chollet *01:18:26*

Absolutely. In the context of the competition, we want to see how much progress we can do with limited resources. But you're entirely right that it's a super interesting open question, what could the biggest model out there actually do on ARC?

We actually also want to make available a private, one-off track where you can submit to us a VM. You can put on it any model you want. You can take one of the largest open source models out there. Fine-tune it, do whatever you want, and just give us an image. There we run it on the H100 for 24 hours or something. You see what you get.

Mike Knoop *01:19:03*

It's worth pointing out that there's two different test sets. There is a public set set that's in the public GitHub repository that anyone can use to train. You can go to an open API call, whatever you'd like to do. There's the private test set, which is the hundred that is actually measuring the state of the art.

It is pretty open-ended and interesting to here folks at least attempt to use the public test set and go try it. Now there is an asterisk on any score that's reported on the public test set, just because it is public. It could have leaked into the training data somewhere.

François Chollet *01:19:52*

This is actually what people are already doing. You can already try to prompt one of the best models, like the latest Gemini or the latest GPT-4 with tasks from the public evaluation set. Again, the problem is that these tasks are available in JSON files on GitHub. These models are also trained on GitHub. So they're actually trained on these tasks.

That kind of creates uncertainty. If they can actually solve some of the tasks, is that because they memorized the answer or not? Maybe you would be better off trying to create your own private, ARC-like very novel test set. Don't make the tasks difficult. Don't make them complex. Make them very obvious to humans, and then see if you can actually make progress on it possible. Make them unique, different, and see how well your GPT-4 or GPT-5 does on them.

Dwarkesh Patel *01:20:25*

There have been tests on whether these models are being overtrained on these benchmarks.

Scale recently did this with GSM8K. They basically replicated the benchmark, but with different questions. Some of the models actually were extremely profit on the benchmark, like Mistral and so forth. Frontier models like Claude and GPT actually did as well on their novel benchmark as they did on the specific questions that were in the existing public benchmark.

It would be relatively optimistic about them just training on the JSON. I was talking with Miles Brundage and he was saying that OpenAI has more private evaluations. Most of these ARC questions are in the JSONs and people are plugging with GPT-4 scaffolding to smart tries to solve the problem. Maybe there's some validation set on the API. If it performs worse than the one we get on the public, maybe allowed the API to access, that means that OpenAI is training on your API calls. You go public with this and show them like, 'oh my god, they've baked your data.'

Mike Knoop *01:21:31*

We do want to evolve the ARC dataset. That is a goal that we want to do. François mentioned that it's not perfect.

François Chollet *01:21:38*

Yeah, ARC is not a perfect benchmark. I made it over four years ago, almost five now. This now is a time before LLMs. We're actually biased a lot since about what potential there their might be. There is some redundancy in the set of tasks, which is of course against the goals of the benchmark. Even the set of tasks is not unique in practice. That's not quite true. Every task is also supposed to be very novel, but in practice, they might not be. They might be structurally similar to something that you might find online somewhere.

So we want to keep iterating and release an ARC 2.0 version later this year. When we do that, we're going to want to make the old private test set available. Maybe we won't be releasing it publicly, but what we could do is just create a test server where you can query, get a score, and submit a solution. Of course we can do whatever frontier model you want there.

Because you actually have to query this API, you're making sure that no one is going to accidentally train on this data. It's unlike the current public ARC data, which is literally on GitHub. There's actually no question about whether the models are trained on it. They are because they train on GitHub.

By going access to requiring this API, we would avoid this issue. Put people who want to try whatever technique they have in mind, using whatever resources they want, that would be a way for them to get an answer.

Dwarkesh Patel *01:23:01*

I wonder what might happen. I'm not sure. One answer is that they come up with a whole new algorithm for AI with some explicit program synthesis. Now we're on a new track. Another is that they did something hacky with the existing models in a way that actually is valid, which reveals that maybe intelligence is more of getting things to the right part of the distribution. Then it can reason.

In that world, that will be interesting. Maybe that'll indicate that you had to do something hacky with current models. As they get better you won't have to do something hacky. I'm also going to be very curious to see if these multimodal models will natively perform much better at ARC-like tasks.

Mike Knoop *01:23:51*

If ARC survives three months from here, we'll up the price. We're about to make a really important moment of contact with reality by blowing up the price, putting a much bigger price pool against it. We're going to see how really quickly if there's a lot of low-hanging fruit ideas.

Again, I think new ideas are needed. Anyone listening might have the idea in their head. I'd encourage everyone to give it a try. As time goes on, that adds strength to the argument that we walled-out in progress and that new ideas are necessary to beat ARC.

François Chollet *01:24:19*

Yeah, that's the point of having a money prize. You attract more people and you get them to try to solve it. If there's an easy way to hack the benchmark, that unveils that the benchmark is flawed. You're going to know about it. In fact, that was the point of the original Kaggle competition for ARC back in 2020. We ran a very strong competition because I had released this dataset and I wanted to know if it was buildable, if you could clear.

There was a small money prize at the time. It was like $10K. This was right around the same time as GPT-3 was released. People of course tried GPT-3 on the public data. It scored zero. When the first contest taught us is that there is no obvious shortcut. Now there's more money. There's going to be more people looking into it. We're going to find out. We're going to see if the benchmark is going to survive.

Let's say we end up with a solution that is not like trying to brute force the space of possible ARC tasks. It's just trained on core knowledge. I don't think it's necessarily going to be in and of itself AGI, but it's probably going to be a huge milestone on the way to AGI. What it represents is the ability to synthesize a problem solving program from just two or three examples. That alone is a new way to program.

It's an entirely new paradigm for software development. You can start programming potentially quite complex programs that will generalize very well. Instead of programming them by coming up with the shape of the program in your mind and then typing it up, you're actually just showing the computer what outputs you want. You let the computer figure it out. That's what is extremely powerful.

(01:26:15) - Possible solutions to ARC Prize

Dwarkesh Patel *01:26:19*

I want to riff a little bit on what kinds of solutions might be possible here, and which you would consider defeating the purpose of ARC or which are valid.

Here's one I'll mention. My friends Ryan and Buck signed up last night because I told them about this. They were like, 'oh, of course LLMs can solve this.'

Mike Knoop *01:26:37*

Good. Thank you for spreading the word.

Dwarkesh Patel *01:26:39*

They were trying to prompt Claude Opus on this and they say they get 27% on the public ARC set. What they did was have other examples of some of the ARC tasks and in context explain the reasoning of why you went from one output to another output and you have the current problem. I think there was also expressing the JSON in a way that is more amenable to the tokenizer.

Another thing was using the code interpreter. Do you think the code interpreter, which keeps getting better so these models get smarter, is just the program synthesis right there? What they were able to do was get the actual output of the calls, the JSON output, through the code interpreter, like "write the Python program that gets the right output here."

Do you think that the program synthesis kind of research you're talking about will just look like using the code interpreter in large language models?

François Chollet *01:27:36*

I think whatever solution we see that will score well is probably going to need to leverage some aspects from deep learning models and LLMs in particular. We've shown already that LLMs can do quite well. That's basically the Jack Cole approach.

We've also shown that pure discrete program search from a small DSL does very well. Before Jack Cole, this was the state of the art. In fact, it's still extremely close to the state of the art and there's no deep learning involved at all in these models.

We have two approaches that have basically no overlap, they work very well. They're very much on two opposite ends of one spectrum. On one hand, you have these extremely large banks of millions of vector programs, but very shallow recombination, simplistic recombination. On the other end, you have very simplistic DSLs, 100-200 primitives, but very deep, very sophisticated program search.

The solution is going to be somewhere in between. The people who are going to be winning the ARC competition and making the more progress towards near-term AGI are going to be those that manage to merge the deep learning paradigm and the discrete program search paradigm into one elegant way.

You asked what would be legitimate and what would be cheating. If you want to add a code interpreter to the system, I think that's great. That's legitimate. The part that would be cheating is trying to anticipate what might be in the test, like brute force the space of possible tasks and train a memorization system on that. You rely on the fact that you're generating so many tasks, millions and millions. Inevitably there's going to be some overlap between what you are generating and what's in the test set.

That's defeating the purpose of the benchmark because then you can just solve it with that and you need to adapt just by fetching a memorized solution. Hopefully ARC will retire that, but on benchmark is perfect. Maybe there's a way to hack it. We're going to get an answer very soon.

Dwarkesh Patel *01:29:41*

Although some amount of fine tuning is valid because they have to use open source language models to compete here and they're natively language. They'd need to be able to think in the ARC-type way.

François Chollet *01:29:58*

Yes. You want to input core knowledge, ARC-like core knowledge, into the model but surely you don't need tens of millions of tasks to do this. Core knowledge is extremely basic.

Dwarkesh Patel *01:30:08*

If you look at some of these ARC-type questions, I actually do think they tell a little bit on things I have seen throughout my life. For example, something Socrates off a wall and Socrates back and you see that pattern. I've played scales piano and I've seen Pong or something.

For example, you see the Pong effect and people's intelligence, as measured on Raven's progressive matrices, increasing on these kinds of questions. It's probably a similar story where since childhood now, we actually see these sorts of patterns in TV and whatever, these spatial patterns.

So I don't think this is core knowledge. This is actually also part of the "fine-tuning" that humans have as they grow up, seeing different kinds of spatial patterns and trying to pattern match to them.

François Chollet *01:30:54*

I would definitely file that under core knowledge. Core knowledge includes basic physics, for instance bouncing or trajectories. That would be included.

But yeah, you're entirely right. The reason why, as a human, you're able to quickly figure out the solution is because you have this set of building blocks, this set of patterns, in your mind that you can recombine.

Dwarkesh Patel *01:31:12*

Is core knowledge required to attain intelligence? For any algorithm you have, does the core knowledge have to be, in some sense, hardcoded? Or can even the core knowledge be learned through intelligence?

François Chollet *01:31:22*

Core knowledge can be learned. In the case of humans, some amount of core knowledge is something that you're born with. We're actually born with a small amount of knowledge about the world we're going to live in. We're not blank slates.

But most core knowledge is acquired through experience. The thing with core knowledge is that it's not going to be acquired in school for instance. It's actually acquired very early in the first 3-4 years of your life. By age four, you have all the core knowledge you're going to need as an adult.

Dwarkesh Patel *01:31:52*

Interesting. On the point itself, I'm super excited to see the open source versions, maybe with a Llama 1TB or something, and what people can score in the competition itself. I'm also excited to see specifically the scaling hypotheses and I'm very curious if you can prompt on the public version of ARC.

You won't be able to submit that to this competition itself but I'd be very curious to see if people can crack that and get ARC working there. Would that update your views on AGI?

Mike Knoop *01:32:23*

It's really going to be motivating. We're going to keep running the contest until somebody goes a reproducible open source version to the public domain. Even if somebody privately beats the ARC and, we're going to still keep the prize money until someone can reproduce it and put the public reproducible version out there.

François Chollet *01:32:37*

Exactly. The goal is to accelerate progress towards AGI. A key part of that is that any meaningful bits of progress need to be shared, need to be public, so everyone can know about it and try to iterate on it. If there's no sharing, there's no progress.

Dwarkesh Patel *01:32:51*

What I'd especially curious about is disaggregating the bets. Can we make an open version of this or is this just possible with scaling? We can test both of them based on the public and the private version.

Mike Knoop *01:33:06*

We're making contact with reality as well with this. We're gonna learn a lot about what the actual limits of the compute are. If someone showed up and said, "hey, here's a closed source model and I'm getting ~50% with it," that would probably update us. We'd think, "okay, perhaps we should increase the amount of compute that we give on the private test set in order to balance."

Some of the decisions initially are somewhat arbitrary in one sense. We don't know what people want. What does progress look like? Both of us are committed to evolving it over time in order to be the best or the closest to perfect as we can get it.

Dwarkesh Patel *01:33:52*

Awesome. Where can people go to learn more about the prize and maybe try their hand at it?

Mike Knoop *01:33:56*

Arcprize.org. It's like now.

Dwarkesh Patel *01:33:57*

It goes live today. One million dollars is on the line, people.

Thank you guys for coming on the podcast. It's super fun to go through all the various intelligence and get a different perspective and also to announce a prize here. This is awesome.

Mike Knoop *01:35:07*

Thank you for helping break the news.

François Chollet *01:33:51*

Thank you for having us.

17 Likes · 3 Restacks

3 Comments

Write a comment...

Top Feed

Colin Bruce · Frequent in and out (Original) · Jun 13
I know nothing about the Arc prize. Really interesting diving into.
Definitely made me think about "scale does not mean we are on the path to intelligence just becoming more and full.
1R · Like · Reply · Report · Hide replies

Jurgen Gravestein · Teaching computers how to talk · Jun 15
By far the most interesting episode so far. Dwarkesh Patel's impressive pre-inviting different voices than you usually have on.
1R · Like · Reply · Report · Hide replies

Deliverance Yeh · Dot Driver · Jun 14
Indeed, but it's a great thinking exercise~
Although the host speaks 250,000 miles an hour
1R · Like · Reply · Report · Hide replies

Ready for more?

Type your email... Subscribe

© 2024 Dwarkesh Patel · Privacy · Terms · Collection notice
Substack is the home for great culture

Start Writing · Get the app