

Will LLMs Get Us to AGI or Are They a Dead End?

Three expert frameworks for thinking about the road to AGI

Mike Mantell

Jan 07, 2026

HeresToSurvivingAI.substack.com

Is human-level AI right around the corner?

I've seen a number of AI safety videos warn us that it is. And they'll cite two of the "Godfathers of AI," Geoffrey Hinton and Yoshua Bengio. Both of whom have issued public warnings that we might soon create something we could lose control of.

But these videos often ignore Yann LeCun, the third "Godfather of AI." These three men are called "Godfathers" because they split the 2018 Turing Award (which is like the Nobel Prize for computing) and helped shape our current understanding of AI.

Unlike Hinton and Bengio, LeCun doesn't think we are on the cusp of human-level AI. He sees LLMs as the current "dominant paradigm," but one that won't get us to human-level intelligence.

Clearly, there isn't a single opinion among the experts.

Will LLMs get us to AGI or not? I wanted to understand the range of perspectives out there, so I interviewed a few prominent thinkers in AI research.

Getting Our Definitions and Acronyms Straight

The question we're exploring is: **can LLMs lead us to AGI, or are they a dead end?**

LLM stands for **Large Language Model**. It's the type of technology ChatGPT, Claude, and other chatbots are based on.

AGI stands for **Artificial General Intelligence**. While there's no singular definition that everyone uses, here's what the Center for AI Safety says:

AGI is an AI that can match or exceed the cognitive versatility and proficiency of a well-educated adult.

Why this question matters

AGI could be a huge deal.

Elon Musk thinks AI will eventually replace all human labor. Sam Altman thinks AI will fix our climate and discover all of physics. Both of these titans and Dario Amodei think it could cause human extinction.

We don't exactly know how big a deal AGI could be. But it seems safe to assume it'll be somewhere between "pretty big" and "the biggest thing in human history."

So if LLMs can get us to AGI, then societal shifts could be right around the corner, in a future iteration of ChatGPT.

But if LLMs aren't the right approach to get us to AGI, then things may fizzle out soon, until the next wave of AI comes along.

Can LLMs get us to AGI? I found three major schools of thought on the matter:

- The Scaling View
- The Real-World Learning View
- The Ecosystem View

Let's dive into each.

Three Views on Whether LLMs Can Lead to AGI

	The Scaling View	The Real-World Learning View	The Ecosystem View
Core Claim	Scale up LLMs to reach AGI	LLMs alone can't reach true intelligence	AGI from multiple technologies
Role of LLMs	Main engine of progress	Limited, need real-world learning	One piece of a larger puzzle
What's Missing?	More compute & data	World models & goals	Unknown future components
Timeline	AGI in 5-10 years	New paradigm required	Highly uncertain timing
Outlook			
Main Risk	Underestimating speed	Overinvesting in the wrong path	False confidence in one approach

Shared Ground Across All Views: - AGI is *plausible* - Timelines are *uncertain*
 - The stakes are *extremely high*

The Scaling View: Meet Nikola

If we keep pouring money into improving our current AI models, we'll get to AGI.

I got on a Zoom call with Nikola Jurkovic. Nikola graduated from Harvard with a degree in AI & Society and now works at METR, an organization that tests leading AI systems' capabilities and risks.

Nikola defines AGI as “something that can automate around 95% of the remote labor that existed in, say, 2022.”

He has a straightforward view on LLMs, saying, “If the question is simply whether LLMs could get us to AGI given enough compute, then I think probably yes. On their current trajectory, I'd expect LLMs to reach AGI in maybe five to ten years.”

Nikola's point about compute being the main bottleneck is a popular one among AI researchers. Dario Amodei also holds this view. He recently told the New York Times, “I think scaling is going to get us [to AGI].”

Nikola's confidence in LLM progress is informed by a popular study that METR conducted. They measured how long of a coding task an AI could complete at a 50% success rate. Nikola explains that “the length of tasks LLMs can do autonomously is roughly doubling every seven months; maybe even faster recently.”

To Nikola, LLMs are already on a reliable path to autonomy. As he puts it: “I don't think we need some fundamentally new paradigm for LLMs to become more autonomous. Given how much they've improved over the last few years, they seem very much on track to become dramatically more autonomous in the coming years.”

Nikola also reflects that progress could speed up even more: “Once LLMs meaningfully speed up AI research, we'll get to AGI pretty quickly after that.”

While AI is all over the internet, Nikola thinks AGI is underhyped.

“A lot of people still don't seriously consider AGI. If the economy truly expected AGI, things would look very different right now. Markets aren't pricing it in.”

He's suggesting that the world could change dramatically: “Taking AGI seriously is an extremely wild idea. The notion that the world could change that quickly is hard for people to accept. A lot of people seem to have a kind of psychological force field against extremely radical ideas.”

I think he's right that it's hard for most people to occupy the Scaling View. Each year of your life has been fairly similar to the year prior. But if we are on the cusp of AGI, then things could get *extremely* different *very* quickly. And that's hard to wrap the mind around.

The Real-World Learning View

While LLMs are impressive, they aren't a pathway to real intelligence. We can only create AGI with an AI system that can learn through directly experiencing the world.

There are a few spins on this view.

Yann LeCun, one of our aforementioned Godfathers of AI, thinks LLMs are “missing something big still.” Unlike Nikola, he doesn’t think AI progress is simply “a question of more infrastructure, more data, more investment, and more development of the current paradigm.” Rather, he thinks we’ll need a new “generation of AI systems.”

To LeCun, the problem with LLMs is that they exist only in text. In an interview with Time Magazine, he explains, “The vast majority of human knowledge is not expressed in text. It’s in the subconscious part of your mind that you learned in the first year of life before you could speak.”

He states it more bluntly in a recent lecture: “Language is a projection of reality, not reality itself.”

He goes on: “There is enormously more information in sensory input—vision, touch, and action—than in all the text ever produced by humans.” And he further states, “To reach human-level AI, we need systems that learn world models from observation, have memory, can plan, and can reason.”

In other words, LeCun thinks LLMs fundamentally lack the potential for intelligence because they don’t have actual knowledge of the world, common sense, or the ability to plan. Talking to ChatGPT *feels* like you’re talking to a person. And its responses *look* like reasoning. But it’s really just predicting text. And the fact that these models still hallucinate is a testament to their inherent limitations.

Richard Sutton, who received the 2024 Turing Award for his work in reinforcement learning, shares LeCun’s skepticism but from a different angle. While LeCun emphasizes that AI would need to understand the world to ever reach AGI, Sutton stresses that AI would need to learn through experience and goals.

Sutton’s core objection is that LLMs learn from text rather than direct experience of the world.

As he put it in one interview, “What we want, to quote Alan Turing, is ‘a machine that can learn from experience,’ where experience is the things that actually happen in your life. You do things, you see what happens, and that’s what you learn from.”

Sutton points out that LLMs learn only through what humans have written about the world, and can’t directly interface with reality.

His other critique is about goals. “Having a goal is the essence of intelligence,” he argues. “Something is intelligent if it can achieve goals.” LLMs, in his view,

“are trying to get by without having a goal or a sense of better or worse. That’s just exactly starting in the wrong place.”

His point is that without goals or real-world feedback, they can’t learn the way animals do, which is through trial, error, and consequences.

If the Real-World Learning View is right, it means that, while LLMs may be useful, they won’t get us to AGI. The technology will eventually plateau, and we’ll need to create another form of AI that can learn through engaging with the world.

The Ecosystem View: Meet Benjamin

AGI will be reached through an ecosystem of AI technologies, and LLMs are just one piece of the puzzle.

To explore this perspective, I met Benjamin Brast-McKie for a drink in San Francisco. Benjamin got his PhD in philosophy and formal logic at Oxford and did his postdoctoral research at MIT. He studies AI reasoning and is working on The Logos Project, which aims to make LLMs more reliable by training and verifying their logical reasoning.

What makes Logos distinctive is its scope: it implements formal verification for highly expressive logical systems. This isn’t the simple propositional logic of introductory courses. Instead, Logos works with rich languages containing tense operators (past and future), modal operators (necessity and possibility), counterfactual conditionals (what-if reasoning), as well as causal, epistemic (knowledge and belief), and normative (obligation and permission) operators. Together, these interacting operators enable precise reasoning about plans in multi-agent systems under conditions of uncertainty.

He takes a more agnostic stance on the role of LLMs, which he says “have already played a substantial role and will likely continue to. But is it the only ingredient? I doubt it.”

Benjamin thinks of AI as an ecosystem. “Rather than fixating on one technology and imagining it scaling infinitely, I think of AI progress as an ecology: an ecosystem with different niches and carrying capacities.” He finds that view useful because in tech, “there are always bottlenecks. One bottleneck creates pressure for new growth, which introduces new constraints. That’s why I don’t like thinking about any single technology as something that will scale infinitely.”

In other words, Benjamin doesn’t think we can simply add more compute to LLMs until we get to AGI. We’ll hit bottlenecks. And once we solve those with new innovations, new bottlenecks will emerge. Which we’ll then solve with further innovations. Etc. But we don’t know which bottlenecks and solutions will arise ahead of time. It’s all impossible to predict.

That’s why Benjamin says, “Making confident predictions about timelines feels misguided. We might spend the next decade elaborating the technologies we already have. Or entirely new actors might appear and change the conversation completely.” To him, progress happens when multiple technologies interact. LLMs may be one of those technologies. But other technologies might not be on the scene yet. Or they could already exist, but have not yet combined in the right way to cause the synergies necessary for a new paradigm to form.

He goes on to clearly state: “It’s not only about which [technologies] haven’t appeared yet. It’s also about which ones are already here, but haven’t yet mixed.” Consider, for instance, how the combination of generative AI and formal verification has been explored in recent years, with impressive results: Apple’s Hilbert system achieved 99.2% accuracy on formal proofs, and Google’s AlphaProof helped solve International Math Olympiad problems. Yet these systems have focused on mathematical reasoning. No one has trained AI systems to reason within the kind of richly expressive formal languages that Logos implements, with their many interacting operators for complex real-world reasoning about plans, causation, and counterfactuals.

This represents a genuinely unexplored combination. Just as AI systems learned to excel at chess and Go - games with explicit, well-defined rules - they promise to be natural at formal reasoning, since logic too is governed by explicit rules for drawing inferences. Adapting the training regimes that produced AlphaGo and AlphaZero to reasoning in expressive logical systems has direct consequences for the challenges of real-world planning and decision-making under uncertainty. After all, we all relate to making and evaluating future or past plans much more than we do to writing math proofs.

To elaborate on his point about technological combinations, Benjamin uses a metaphor about physics. “Physics only became what we now recognize once geometry, algebra, and analysis came together. Those ingredients existed separately for a long time. But once they combined in the right way, you could suddenly articulate a mechanics.” The historical parallels are instructive. Descartes invented analytic geometry by combining algebra with spatial reasoning. Newton and Leibniz developed calculus by fusing infinitesimal methods with rigor. Each synthesis created something genuinely new from ingredients that had long existed separately.

Perhaps most exciting are the connections between generative AI and the recursive semantics guiding the development of the Logos. All languages require interpretation: we need to construct models that give meaning to the operators that we use. This is precisely what we do when we imagine scenarios, consider possibilities, or trace out consequences of our actions. Generative AI shows remarkable ability to build rich representations from context. The synergy could run in both directions: AI helps construct the formal models from natural language descriptions, while formal methods verify and constrain AI reasoning. You don’t have to trust the AI if it can check its own reasoning with mathematical certainty.

If The Ecosystem View is right, then the question of whether LLMs will get us to AGI isn't the right question to ask. AGI may well be on the horizon, and LLMs are no doubt a major technology that will help us get there. The more interesting question is what ingredients are missing, and which existing ingredients haven't yet combined in the right way.

Comparing Each View

The real disagreement isn't whether AGI is possible, but what path will get us there, and how predictable that path is.

The Scaling View assumes we can see the path ahead. We'll get to AGI with more compute, better architectures, and incremental improvements.

The Real-World Learning View argues we're on the wrong path entirely and need a fundamental paradigm shift.

The Ecosystem View suggests the path is fundamentally unpredictable; we can't know which technologies will matter until they combine in unexpected ways.

These differences in opinion have stakes. They suggest how much time we have to prepare for transformative AI. And they may imply where billions of dollars get invested, and how to think about safety.

Where I've Landed on This Question

First and foremost, I've come to believe that it's important to hold uncertainty on this topic.

The truth is, nobody knows how this technology is going to happen. Or when. There are very smart, confident people on all sides of the issue. But at the end of the day, everyone is making a guess.

However, what struck me most is that all the experts do have one key belief in common:

AGI is coming.

And fairly soon.

LeCun, our famous LLM skeptic, believes AGI might arrive within 10 years. And Sutton, our other famous skeptic, implies human-level AI could arrive in the 2030s.

To make that crystal clear: even those who are most skeptical of LLMs still think AGI could be here by 2040!

So, in a sense, the question of whether LLMs get us to AGI isn't all that relevant.

In fact, few AI experts think AGI is more than 25 years out.

In 2023, surveyors asked 2,778 AI researchers when they thought AGI would arrive. The average prediction was a 50% chance that AGI arrives by 2047.

21 years from today.

If The Scaling View is wrong, and AGI isn't just around the corner, it'll be a blessing. We'll have gotten a dress rehearsal for handling AGI. We can learn from how we handled this moment and do better when the real thing arrives.

To be honest, I don't think we're handling AI safety very well so far. Companies are careening forward at blazing speeds. Profit and speed have far outweighed safety.

In my view, the stakes could not be higher. And the timeline is uncertain. So I think we should treat the Scaling View with real sincerity.

I'll close by paraphrasing a thought experiment from AI research pioneer Stuart Russell.

Imagine we receive a message from an alien species that says:

"We are an alien species with technology extraordinarily more advanced than humanity's. We will visit your planet sometime in the next 3-50 years. Please prepare yourself."

This could be the situation we are in. And if AGI does arrive, I hope we greet that day feeling prepared emotionally, ethically, and politically.