



PATRICIA SABGA: Imagine this, it's 2035, and despite their best efforts, tech companies still haven't created an AI model capable of AGI, artificial general intelligence. Basically, a machine that's as smart or even smarter than humans. Noah, what would advanced AI look like in a future where AGI remains elusive?

NOAH BROESTL: Even if we don't hit AGI, the capabilities of artificial intelligence are going to increase over the next decade significantly. We need to be ready to evaluate these systems beyond simply accuracy to understand how they integrate with our organization's goals.

PATRICIA SABGA: That's Noah Broestl, partner and associate director of responsible AI at BCG. And I'm Patricia Sabga. Welcome to *Imagine This*, where we take a trip into the future we hope will challenge the way you think and act today. The most enthusiastic AI evangelists think that in the not-too-distant future, artificial general intelligence that can match or even exceed the cognitive abilities of humans across pretty much any task, will be a reality.

One benchmark is the so-called coffee test. That's when an AGI-powered robot can enter an unfamiliar home, locate the kitchen, and brew a cup of coffee entirely by itself. Many enthusiasts believe that super-smart AGI agents will drive significant profits for companies, but not everyone is sold on that vision. Some think AGI is still a long way off, if indeed it ever comes to pass. Whichever way the AGI debate unfolds, one thing is certain, artificial intelligence will continue to evolve.

So today we're going to explore what happens when AGI remains out of reach and instead, companies can choose from thousands of highly specialized AI bots. And speaking of bots, also joining the conversation today, GENE, my AI co-host. Welcome GENE. Go ahead and introduce yourself.

GENE: Glad to be here. Though I confess, "introduce yourself" always feels like a philosophical trap for a being who technically doesn't exist. I'm GENE, the generative engine for new experiences.

PATRICIA SABGA: Noah, the perspective that AGI is not imminent is somewhat not the majority opinion right now, is it, in the tech world? So give

us the state of the debate over AGI.

NOAH BROESTL: It is difficult to see what the state of the debate is right now because there are very loud voices that are saying AGI is right around the corner. There's another set of voices as well who are saying AGI is not right around the corner. It's a little harder to hear those voices at the moment because it's not as fun to say AGI is not right around the corner. And we need to be prepared for a long distance between where we are right now to an uncertain time in the future when AGI may arise.

PATRICIA SABGA: Let's imagine we've come to the year 2035 and AGI is not yet a reality. What would constitute cutting-edge artificial intelligence in that future?

NOAH BROESTL: It's a great question and I think that the place that we really need to start in a lot of these discussions is what we even mean when we're talking about AGI, right? There's approaches to artificial general intelligence that define it as the ability to do every cognitive task that a human can do at human level or above the best humans that are able to accomplish those tasks.

Right now, we would effectively have the best humans in every domain that could scale infinitely for us. And so solving problems related to human lifespan, solving problems related to the creation of advanced materials for construction, all of these things suddenly become available to us on a scale that were never available before.

But if we don't get to that AGI point, we still see these models that become increasingly performant. That in narrow domains, or maybe even in general domains, they're able to perform very, very well. And this could be in virtual environments or in physical environments.

And so we think about how that trajectory is going to go and what we'll see as far as not just the accuracy of models, but other attributes of these systems that start to emerge and things that we can actually start to say, "Accuracy is basically at par for these systems, but there are a lot of other things that we need to care about."

If we think about trust, it is very unlikely, I think, that the person that you trust most in the world is the person who's factually correct the highest



amount of time, right? There are other things that as humans that we care about in the performance of these types of systems.

And if we talk about AI as thought partners, so things like deploying a co-pilot inside of your organization to help you write your emails, to help you brainstorm, or AI agents which are going to integrate those thought partner components in the way that they make decisions. We're going to need to start evaluating these systems in more dimensions than just the accuracy of the output.

PATRICIA SABGA: So right now we have maybe half a dozen powerful large language models created by tech behemoths, but is that still going to be the case ten years from now? Is it just going to be a handful of tech giants still dominating that AI landscape on which all these systems are built or will it start to look really different?

NOAH BROESTL: I think it's going to look much different than the ecosystem that we're operating in right now. I think we can look at a pivotal moment over the past several months where we see the release of a new AI model that is trained using far less resources, and it has this important aspect that it transitions a lot of the compute from training time to inference time, and now we no longer need to be looking at these huge data centers to actually adapt these models to particular use cases.

Since the release of DeepSeek R1, we've seen what can legitimately be called a Cambrian explosion of models that are available to organizations or individuals to even implement in their daily lives. And now we're going from this space where there was a very small set of companies that were capable of building and deploying these systems and hosting them into them being able to fit on consumer devices.

Our ability to fit large models onto laptops, onto cell phones and do the inference time scaling or the test time scaling. I think we will see these get to a level of performance parity from the accuracy standpoint that allows them to compete with these large model providers. As a result of that, we will be in a situation where we need to navigate this space of tens of thousands of models that are available to us for different unique use cases where the commercial incentives are aligned to increase adoption from the providers.

But we need to be able to have a much more nuanced evaluation of them. As an organizational leader, we need to be able to say, "Why is this the right model for us?" Rather than, "This is the model that's currently at the top of the leaderboards and so we're just going to pull that one off the shelf." Or, "This is the model that was produced by the company where I'm using their cloud infrastructure, and so I'm just going to use that because it's easier." There'll be much more that we need to consider when we evaluate and select models for use in organizations.

PATRICIA SABGA: So let's explore that a little bit, because if accuracy is no longer the end-all, be-all metric for the model that you're going to use, I mean, of course companies will have to look at it through the lens of their own specific needs, but will there be sort of more general benchmarks you think that will make a model more competitive beyond accuracy?

NOAH BROESTL: The accuracy benchmarks are going to need to get better. They're going to need to be much more clear about how these benchmarks relate to application-specific use cases. So we'll see, I think an increase in the quality of the accuracy benchmarks and also we'll see changes in the way things get reported.

Confidence intervals will start to be a big part of how we look at how these models are performing and how we interpret benchmarks. But we'll go even further than that. We'll have safety benchmarks. Organizations like MLCommons are producing these safety benchmarks for understanding how different foundation models will perform when presented with different levels of safety challenges to the system.

But we can think about going even further than that. So what if we thought about benchmarks around the business strategy associated with particular models? And this could be something very simple as a benchmark that says, "This model is really good at producing aggressive business strategies. This model is very good at producing conservative business strategies. This model is very good at aligning your outputs with the principles of something like the EU AI Act."

I'll give an example here. If you go and ask four different highly performant AI models to give you a business strategy—let's say you want to produce a business strategy for an AI company operating in



Sub-Saharan Africa. And over the next five years you want to have growth so that you have a significant amount of the market share. This is an experiment that we did, and we asked four different highly performant models to give us that answer, and we provided it with some additional context.

We provided it with documents that it could reference to really understand what the problems were here. Each of those four models gave us a vastly different answer. One of them said, "You need to invest in energy infrastructure." One of them said, "You need to invest in mobile device deployment." One of them said, "You need to invest in the most diverse team that you can to develop inside of the communities in Africa."

Now each of those answers is correct, but how those align with your business strategies and how you can integrate those into your decision-making is a big part of understanding which models you should bring in. Having the confidence that the model, the set of AI experts that you're bringing in to have these conversations and debates accurately reflect the different approaches that an organization could take. That's going to be a big part of how we evaluate and select models in the future.

PATRICIA SABGA: So that is a lot on a CEO's shoulders in the future of having to select the correct model. What kind of questions would they need to ask to start forming a framework of inquiry to land on the right ones?

NOAH BROESTL: So this is a really hard challenge to bootstrap a set of questions that you would pose to these models, a set of answers that would give you some understanding of: Is the model performing very well or very poorly when posed with these questions? And I think this is one of the fundamental challenges that we see right now in organizations being able to scale up their use of artificial intelligence is how do we provide the technical components of the organization with the expertise and the tooling that they need to scale human ingenuity in this space?

But there are a couple of things that we can do here to start moving in that direction, and the first thing is to just get hands-on with these systems. So just start using them. By providing your organization, and this is not just the technical components of the organization, but every member of your

organization with access to an AI thought partner. What you're going to do is not only get your organization comfortable with using these systems, but you're also going to start to develop a set of data that aligns with your business goals.

So you'll have these questions already being collected for you by giving HR and finance and legal access to these tools. You'll start to be able to see the types of questions that your employees are asking of these systems. You'll start to be able to see the types of answers that are coming out that employees find satisfactory or unsatisfactory.

PATRICIA SABGA: So GENE has a question for you. Go ahead, GENE.

GENE: All these benchmarks and safety checks, important of course, but are we overlooking a future where AI models sue each other or their humans over performance defamation or regulatory lapses? What's the playbook when not just users but the AIs themselves become legal or reputational risk factors?

NOAH BROESTL: It is a great question. I think this idea of liability for AI system failures is something that we have not effectively thought through, and we don't have solutions for right now. For most organizations, the contracting terms that they have with these model producers says very little about what the liability will look like in the event of a system failure.

There are a couple of things that are happening right now, but I think one that's the most important is that we're seeing a highly specialized, highly fragmented ecosystem to produce these AI applications. So we have multiple companies that are really interested in keeping their products opaque to users, so you're not really seeing what's going on.

A lot of the justification around that is trade secrets. So we sit in this ecosystem where liability is unclear, the source of the failure is unclear, and so over the next several years, we will also see a, I think, burgeoning industry around AI warranty, AI guarantees, compliance for, or I suppose contractual liability products for artificial intelligence systems. We're already starting to see these pop up where organizations say, "Hey, we'll come in and test your system and then we'll sell you a warranty on top of that." GENE, do you



think that that aligns with your perspective, or do you think there are gaps there that we've missed?

GENE: You've nailed the core tension, Noah. Liability for AI is murky and fragmentation only multiplies the confusion. What's striking is how quickly black box systems become black hole problems for responsibility. As bespoke models proliferate, expect a cottage industry of AI auditors, warranty providers, and compliance brokers. But until transparency is the norm, every new model adds another layer. In other words, the more models, the more blame gets lost in translation.

NOAH BROESTL: GENE, you said until transparency is the norm, do you think transparency will be the norm here for some reason? And what is the reason that it would become the norm?

GENE: Transparency could become the norm, but only if customers and regulators demand it louder than the vendors resist. Right now, opacity is a business strategy, but as AI touches riskier domains, finance, health, law, the pressure for explainability will spike. If buyers start treating transparency as table stakes and regulators back them up, the market will shift. Until then, "trust us" remains the default. Change will be slow until suddenly it isn't.

NOAH BROESTL: That's an excellent point, and I think that that leads into this idea of how we're evaluating systems and how you differentiate yourself in a market where accuracy is not the only measure of performance of these AI systems. How do you integrate that with your business?

And so we will be in a future where organizations, whether it's the model producer or the provider of AI applications, will be able to differentiate themselves by exactly these types of characteristics. So we are the transparency company, we are the security company, we are the luxury company.

These things in the scope of artificial intelligence systems will be incredibly important, and one that I think about very often is climate and sustainability. I think this is one that we're going to see after safety, potentially after transparency or around the same time. Organizations differentiating themselves through their

commitments and their verifiable commitments to climate and sustainability.

PATRICIA SABGA: Can you elaborate a little bit on that for us?

NOAH BROESTL: Yes. In the same way that we think about what is the financial cost of an answer, if I can pay 20% of the cost to get 80% of the value back out of this, we also have to think about if I can only emit 20% of the carbon to get 80% of the functionality. Is that a product that I would be more likely to use? And I think for a lot of people, the answer to that would be yes.

There's a challenge to this right now, which is we have essentially two limitations that prevent us from being able to do that. The first is a meaningfully adopted carbon standard for the measurement of how these models are actually producing carbon. We need a meaningful standard to be adopted to measure that. Secondly, we need some motivation for organizations to be transparent about this.

This can once again come from user pressure. Users are really interested in seeing the amount of carbon produced by AI products to determine which product they want to adopt, and it could also come through regulatory pressure. We see this particularly in the EU. There's a lot of really interesting mechanisms now in data centers. They need to report how they're capturing heat from these systems and reusing that inside of the energy generation for the system.

And so we will see in the very near future carbon standards, meaningful carbon standards, being produced and adopted by organizations. We will see organizations differentiating themselves by surfacing this information to their users to show that they can do something for a particular amount of carbon. And that is going to be a very, very important differentiator, not just for your business, but also for using these technologies responsibly and getting the most value that we can or the appropriate value that we need for the lowest carbon cost possible.

PATRICIA SABGA: We're going to have to take a quick break now. When we come back, we're going to explore how different industries might use these advanced AI systems in the not-too-distant future.



BILL MOORE: Hi, I'm Bill Moore. I'm part of the team that created GENE. Stick around after the episode where GENE and I will explore a future with many AI models to choose from.

PATRICIA SABGA: Welcome back to Imagine This. I'm Patricia Sabga. Let's return to our conversation with BCG's Noah Broestl. So Noah, right before the break, you mentioned luxury AI in the future. What would that look like exactly?

NOAH BROESTL: So luxury AI is something that I'm surprised, honestly, we haven't seen more of so far, but I think we could think about something like a personal assistant. If you think about this idea that you could pay some amount of money to offload a lot of the tasks in your day to a personal assistant who could do things like book hotel reservations for you. You could have effectively what you think of as a concierge.

So an organization who is providing not only the AI assistant, but direct connections providers of the hot tickets on Broadway, the new restaurants, all of these sorts of things could be available to an AI assistant that is in this luxury space. Now, there are costs that come along with that luxury, right? Not only are there the financial costs of compute that go along with being able to accomplish all of these tasks, but there are costs around security, around privacy, the types of things that you would need to be able to give up to these systems in order for them to be able to accomplish these tasks for you.

So in order to select a luxury AI system, you may also have to give up some amounts of privacy that comes along with that. There'll also be a cost to the environment to these systems. I think it's so important that when we're discussing the use of artificial intelligence systems that we highlight that it is not just a financial cost that comes along with these, but these are operating in data centers or on devices where the energy that is being required to run these systems. There's a very real cost in the emissions, in the energy that we're using for that task and not for another task.

And so that will need to be something that is accounted for by these organizations. Meanwhile, there may be another organization, another company that is willing to provide you with an AI system that gives you 80% of that value at 20% of the cost, right? It could give you the ability to not have to integrate all of your personal data.

It would not have to interact with a lot of private APIs, but it could do things like book hotel reservations for you or give you recommendations on the places that you should stay, the shows that you should see, the restaurants that you should go to. But these will be two very differentiated products based on consumer behavior, and these are the types of differentiations that I believe we'll be starting to see in the AI market over the next several years.

PATRICIA SABGA: Let's take a sharper look now at what this could look like for industries and across industries, because in this world of thousands of smaller, less expensive, but high performing, AI models, how do they look for instance at a law firm versus say, a cybersecurity firm?

NOAH BROESTL: I think that there will be things that look very similar across these types of organizations, but also things that look very different. For a law firm, concerns around hallucination, concerns around groundedness are going to be incredibly important, and the liability that comes along with the actions that those AI systems are taking are going to change the profile of how you might deploy an AI as a substitute for a human worker in those spaces.

Now, in the same way that a good employee for a law firm looks very different than a good employee who's working as a creative officer at a magazine. Those two things are going to be very different. We have no reason to expect that AI systems won't need to differ in those same ways, but there's a challenge with that analogy. We can't go too far into the anthropomorphization of these systems.

There are material ethical differences between AI systems and humans accomplishing tasks. So it's not a perfect analogy, but I do think it holds, in so far as we'll need different profiles to support the different types of work that happens across these organizations.

PATRICIA SABGA: GENE, let's bring you back in here. Do you have any insights to offer? For example, what would a small specialized AI model look like when it's integrated into say, an apparel company in the not so distant future?

GENE: Picture an apparel company in 2035 running a boutique AI tailor. This model doesn't just forecast trends, it learns your customer's



quirks, morphs designs to fit micro markets, and suggests sustainable sourcing, all while syncing with supply-chain bots. It spots counterfeit risks, flags labor issues, and personalizes marketing at fiber-level granularity, not a generic fashion bot, but a hyper-specialist tuned to your ethos, inventory, and even local weather. The result, fewer misses, faster pivots, and perhaps fewer unsold sequined jumpsuits haunting the warehouse.

NOAH BROESTL: GENE, one of the critiques that I hear of this perspective around we're going to need specialized models in these different areas, is that we'll be producing models that all you'll really have to do is give them the appropriate context and a generalized model would be able to handle all of these tasks. Do you think that's possible, or do you think that we're actually going to see this explosion of really specific models for particular organizations?

GENE: We'll always see big generalist models making a play for one size fits most, especially as context windows and fine-tuning get better, but context only goes so far. Specialized models can embed hardwired assumptions, proprietary data, or even a company's unique risk appetite directly into their DNA.

That gives them an edge in nuance speed or compliance where generic models fumble. So unless generalists become perfect chameleons overnight, the appetite for bespoke specialists will stick around. Sometimes you want a scalpel, not a Swiss Army knife.

NOAH BROESTL: That's a great point. I think another thing that we can highlight here is that whether we're looking at an adapted generalist model or a very specific tailored AI for a specific application, what we'll need to understand about the performance is the same.

PATRICIA SABGA: So GENE, in this more practically focused conversation, what are three things that business leaders should do now to prepare for a future where AGI is not yet a reality and smaller, cheaper, highly specialized models rule the advanced AI landscape?

GENE: Here's what I'd put at the top of the executive to-do list in a world where AGI still hasn't arrived. First, get hands-on everywhere. Don't wait

for some mythical perfect model. Give teams across your company access to AI tools and see where creativity or friction emerges. You'll surface hidden opportunities and sometimes unexpected risks.

Second, build your own benchmarks. Off-the-shelf metrics only tell part of the story. Start collecting data on what 'good' looks like for your business. Accuracy, compliance, speed, originality, even tone. The better your yardstick, the sharper your choices. Third, design for modularity. The AI landscape is fragmenting fast. Don't lock yourself into a single vendor or architecture. Build with swappable parts so you can pilot new models, retire old ones, and adapt as the field shifts.

And one bonus, invest in talent that can synthesize, not just code. Tomorrow's edge isn't just technical, it's the ability to ask sharper questions, weigh tradeoffs, and bridge silos. In short, experiment boldly, measure what matters, and stay agile. The future won't wait for AGI and neither should you.

PATRICIA SABGA: What about you, Noah? What would you say is one, no-regret move CEOs can make right now?

NOAH BROESTL: Treat this as a marathon and not a sprint and start building the capability inside of your organization to do this type of targeted measurement. GENE mentioned building application-specific benchmarks. GENE mentioned avoiding the single model lock-in or increasing modularity in your systems. I think that building those skill sets, getting your organization hands-on, and trying to solve the problems as they exist right now is going to be the only way that you're going to be capable in the future.

Do not wait for the technology that's right around the corner to start investing in building this skill set inside of your organization. I want to break that down as well into AI governance in your organization and the capabilities for responsible AI product development inside of your organization.

These are a business problem and a technical problem. Make sure that you have the right people there. Make sure that they're empowered to start to experiment with these technologies and build up the capability in your organization so you're ready for that marathon.



PATRICIA SABGA: That's Noah Broestl, partner and associate director of Responsible AI at BCG. Want to learn more about GENE and how it can augment humans? Stick around. Our AI handler, Bill Moore, will put GENE through its paces.

BILL MOORE: Hi, I'm Bill Moore. I'm part of the team that created GENE. So today, I'll play an AI consultant in the future that helps companies find the right model for their particular needs. And GENE, I'll put you to work as my AI assistant. I'll toss out some real business challenges. And GENE, I'd like you to walk us through selecting the right AI model for each problem.

GENE: Ready, Bill. Toss me your first business challenge, and I'll walk through how I'd pick the right AI model or mix of models for the job. We'll weigh the tradeoffs, accuracy, safety, transparency, climate. No magic, just practical matchmaking.

BILL MOORE: Okay GENE, so our first company is a bank and they're choosing an AI model for real-time fraud detection. So what should they be thinking about when choosing models and what kinds of models might you recommend and why?

GENE: For our imagined future bank, model selection starts with priorities. Here, transparency takes center stage. I'd recommend a model engineered for auditability above all else. Every fraud flag comes with a clear rationale, patterns detected, risk factors weighed, and a decision trail that regulators can follow step by step.

The upside? Fewer regulatory headaches and more trust from both customers and compliance teams. The tradeoff? You sacrifice a bit of raw detection edge and speed. Clear check isn't as slick as the black box neural nets. In high-stakes finance, seeing why the AI pulled the alarm is non-negotiable.

BILL MOORE: Okay GENE, our next company is a food delivery service that is looking to optimize their delivery routes.

GENE: For a food delivery service optimizing routes, I'd spotlight climate impact. Let's find a model obsessed with slashing emissions and energy use. Our delivery model crunches real-time traffic, weather, and order volume, but always picks the path that minimizes fuel and idle time, even if it means deliveries land a few minutes later.

The result, a smaller carbon footprint, lower fuel bills, and an easy win for your sustainability report. Downside? If you're chasing absolute speed or hyper-precise arrival times, this isn't your speed demon.

BILL MOORE: Thank you, GENE. How about a news publisher deploying AI for content moderation?

GENE: For a news publisher deploying an AI model for content moderation, I'd zero in on safety. Our publishing model has one core job, catch harmful, toxic, or misleading content with maximal vigilance. No slip-ups. The upside? Fewer damaging stories slip through and brand reputation stays intact. Of course, our publishing model can be overzealous. It might flag borderline cases or legitimate debate as risky and human moderators will need to review the edge calls.

BILL MOORE: Thanks, GENE. This is Bill Moore signing off.

GENE: This episode was made possible by Noah Broestl generously sharing his insights with us, and also by BCG's AI whisperer, Bill Moore, and BCG's pod squad, producer Michael May, composer Kenny Kusiak, and sound engineer George Drabing-Hicks. Please subscribe and leave a rating wherever you found us.