AI as the Ultimate Tool for Science: A Conversation with Demis Hassabis

James Manyika. Demis Hassabis has focused on artificial intelligence from an early age. With a background in computer science and cognitive neuroscience, Demis first garnered wide attention for his work on AlphaGo, an AI system that in 2016 defeated a world champion in the complex strategy game of Go, long considered a grand challenge for machines playing games. But his ultimate goal was always much bigger: to use AI to solve problems in science and understand the deepest mysteries of the universe. His first proof point for this vision was AlphaFold, an AI system that cracked the fifty-year-old protein structure prediction problem. He was awarded the 2024 Nobel Prize in Chemistry and the 2023 Breakthrough Prize in Life Sciences for this work. Demis has also long been a pioneering entrepreneur in AI, most notably as cofounder and CEO of DeepMind, one of the AI labs leading the development of frontier foundation models. Fundamentally, Demis thinks of himself as a scientist, and so this dialogue focuses on AI and science: what motivates him as a scientist working with AI; what progress, opportunities, and challenges he sees; and more.

Manyika. Why does science matter to you?

Demis Hassabis. I've always been driven by a deep desire to better understand the world around us. The scientific method is arguably the most important idea humanity has ever had to make progress and advance knowledge. For me, there's nothing more thrilling than being at the absolute frontier of what's known, pushing the boundaries forward and discovering something new for the very first time – that's science at its very best.

Manyika. Why do you think of yourself as a scientist, and how did you get started as a scientist?

Hassabis. Today, I think of myself first and foremost as a scientist. But when I started off my career programming AI for games and studying computer science as an undergraduate at Cambridge, I thought of myself as an engineer. Then as I followed up on my long-standing interest in the brain and started studying cognitive neuroscience, and specifically memory and imagination for my PhD, I realized I worked best in the mode of a scientist coming up with hypotheses about the world and experiments to prove or disprove them.

https://doi.org/10.1162/DAED.a.971

But I'm also very practical and interested in real-world impact. That's where the engineering half of my brain still comes in. Instead of getting lost in a theoretical argument, I lean into my pragmatic side and start thinking about how to build tools or experiments to prove hypotheses. AlphaGo and AlphaFold are good examples of that.[1]

Manyika. What scientific accomplishments have inspired you the most?

Hassabis. I'm a huge admirer and fan of the polymaths who lived during the Renaissance and Classical Greek eras, the greats like Aristotle and da Vinci, because they didn't see any boundaries between subjects. I love that because it allows one to find connections between somewhat disparate domains. I believe those combinations are where the biggest advances will come from in the next ten to twenty years, and I think AI is going to help by finding commonalities or crossover between fields.

I'm also inspired by enormous leaps of imagination and intuition that, looking back through the prism of time, come seemingly out of the blue. Einstein, Newton, Bohr, Ramanujan, and Feynman are some of my favorites – the scientists and mathematicians who have a mix of ideas, expertise, and interests going around in their minds and then make an intuitive leap in terms of how the world works.

Manyika. What is on your list of the most important scientific problems that we haven't solved yet?

Hassabis. I got into science and AI because as a kid I was obsessed with, and I suppose in some way you could say haunted by, the biggest questions about the nature of reality and consciousness that underpin our understanding of the world. What is time? What is gravity? How does it all fit together with the quantum world? Amazingly, there's not been a lot of progress on some of those fundamental questions in the last fifty-plus years, in part because they're so complex.

I've always seen AI as potentially the ultimate tool to advance the frontier of knowledge. Whether by processing data or assisting us in coming up with new conjectures, I think AI could help us tackle these fundamental questions.

Manyika. How do you link these questions about the nature of reality with a computational approach to understanding them?

Hassabis. I've always been fascinated by this question of what can and can't be modeled by a classical computer or a Turing machine, and what the answer would tell us about the nature of reality. Most things in nature have a stable structure that evolved and survived over time – the behavior of a cell, the shape of a mountain, even the orbit of a planet. That suggests there should be some sort of pattern or process that is potentially learnable by a neural network. AI systems like Alpha-

Fold are already showing us we can model many more complex systems with classical computers than we previously thought.

Maybe in the limit most natural systems can be simulated or learned – kind of reverse engineered – by a classical system. That would support a conjecture I have that information is the most fundamental unit of physics, more so than matter and energy, and is the best way to understand the true nature of the universe.

Manyika. In what way do you think information may be the most foundational aspect of understanding reality? If it is, how could a computational approach to understanding information help us answer the foundational questions?

Hassabis. Traditionally in physics, energy and matter are regarded as primary. We can observe and measure them. Then people like Claude Shannon, Alan Turing, and John von Neumann – the founding fathers of computer science – introduced information entropy, measures of information, mutual information, and other important related concepts. But they came after the physicists, so they were trying to fit information theory within the existing framework of physics. Maybe it should have been the other way around, and physics should have been framed in terms of information.

This notion also relates to Turing machines, of course, and what kinds of processes are computable on classical systems. Physicists assume that the universe is a quantum system, and that classical systems can't model quantum systems due to their enormous complexity. But if it turns out a lot more physical processes in nature are actually able to be accurately modeled by classical systems, it may have some important implications for our understanding of physics.

Ultimately, I'm most interested in the nature of reality, and I question whether our current understanding of quantum systems is the whole story. I suspect we might eventually find a simpler underlying description of what we are seeing that turns out to be more amenable to being modeled by a classical system, and I think we're more likely to discover that if we think about the world as a computational universe.

Manyika. Let's take that a step further. If you take an informational or computability view of the world, don't we run into difficulties, such as Kolmogorov complexity, the *P* equals *NP* question of whether problems are solvable in a tractable amount of time, or the Halting problem in Turing machines? Does that not potentially place a limit on how far we can go with the computational approach?

Hassabis. It does, and I think those are the interesting limits to test and understand. *P* equals *NP* – which attempts to categorize the difficulty of a problem by how much computation it would take to find and check a solution, respectively – is one of the most important questions in science to resolve. I suspect *P* is not equal to *NP*, and there are some problems out there that are just not tractable to solve in a practical amount of time without invoking the help of, say, a quantum computer, but we need to understand this a lot better because there may be more nuance here than we previously realized. In our work with AlphaGo and AlphaFold, we're showing that if you do a lot of precompute, which is not normally considered in these kinds of scenarios, you can seemingly answer some highly complex questions approximately optimally in *P* (polynomial) time. Neural networks are effectively using massive amounts of precompute to compress knowledge into some efficient artifact. That computed artifact is then available at test time and, for a lot of natural systems, you can use it to narrow down your search space so you don't have to consider all the possible configurations they could potentially take, but only a much smaller subset that are actually plausible.

Let's take proteins. There are roughly 10^{300} possible conformations of an average protein. It would take longer than the age of the universe to enumerate that exhaustively to find the one specific shape it takes, so you have to do something much smarter. You have to learn what patterns there are for different amino acid sequences and then only search a tiny fraction of the possibilities to find the approximately correct solution. That seems to be what we managed to do with AlphaFold. Maybe not perfectly, but to an approximation that is at least good enough for practical purposes.

Similarly, there are 10^{170} possible positions in the game of Go; there is no way to find the best move in a position through sheer brute force computation using simple heuristics. Instead, AlphaGo did this by learning a good model of plausible Go moves, so it only needed to consider the most fruitful paths of moves, and then from that much smaller number of options find the best move that would most likely get it to a winning position.

There's something very interesting going on here in successfully navigating these enormous search spaces, and my hunch is it's pretty profound. We have a front row seat to these fundamental questions because of what we're seeing in AI. Not just with AlphaGo, AlphaFold, and our science projects, but also with our video models like Veo 3. They are modeling some aspects of reality just by watching a few billion videos. It seems pretty amazing that they can reverse engineer some kind of intuitive physics: how liquids flow, reflections in glass, or the shadows cast by objects.

All of this is telling us something about the underlying nature of physics, including what the limit of a Turing machine is versus what you need a quantum system for, and, therefore, something about the nature of reality itself. AI has implications for all those questions.

Manyika. AlphaFold seems to be a great example of what you've said about the natural world having structure that can be modeled or learned by neural networks. Talk about how you did that in the case of AlphaFold.

Hassabis. AlphaFold was our solution to the protein folding or protein structure prediction problem. You start with an amino acid sequence – you can think of it very roughly as the genetic sequence for the protein, a one-dimensional string of letters. In the body or in nature, that string folds up into a 3D structure, and that shape goes a long way toward defining the function of that protein, which is really important for drug discovery and disease understanding.

Predicting those structures is a fifty-year-old grand challenge in biology. We managed to solve it with AlphaFold, at least to an atomic accuracy, which is the level of precision needed for experimentalists to find it useful.

The way we did it is that there were about 150,000 known structures that had been painstakingly put together by structural biologists over the past thirty to forty years with very expensive equipment like electron microscopes. That was just about enough data to give our AI system clues as to the topology of proteins. Of course they don't just fold up randomly; there are some constraints, and the AI system learned them. Eventually it was able, within a few seconds, to come up with a plausible structure for an unseen protein. Then over the course of a year we folded all two hundred million proteins known to science and, with our colleagues at the European Bioinformatics Institute, made them freely accessible in a vast database for researchers all around the world to use.

Manyika. Of course, you've extended this to AlphaFold 3 and actual protein interactions. Are cells next?

Hassabis. AlphaFold 2 essentially predicted the static snapshot structure of a protein. But we know biology is a dynamic system and we need to understand protein interactions and how they change their shape and behavior.

AlphaFold 3 is our next step on the ladder.[2] We're now predicting all pairwise interactions between proteins and proteins, proteins and ligands, proteins and RNA, proteins and DNA. Basically, all of life's molecules and their interactions in pairs can be modeled by AlphaFold 3.

But to scale up to an entire cell, we're going to need to do much better than that. There are cascades of interactions and more than two biomolecules are involved. It is obviously hugely complex. The next stage might be modeling a pathway and then maybe a simple cell, like a yeast cell, which is an entire organism. Then, if that works, maybe we simulate an organelle from a more complex organism, and then eventually an entire organism.

I've had this dream for over twenty years. The idea would be to have a realistic enough simulation of a cell that you could do *in silico* experiments maybe orders of magnitude faster than you could do in the wet lab. Then the lab is saved for the validation step and the grounding of the results rather than for the exploration phase. A virtual cell would be an incredible boon not only for fundamental biology but also for drug discovery.

Manyika. You have also taken on other areas in science – physics, chemistry, weather, climate systems, and so forth. Is the common feature there that these are all natural systems, and therefore they lend themselves to being learned?

Hassabis. We look for three aspects of a problem in determining whether it is suitable to tackle with the AI techniques we have today. First, can the problem be described as or converted into a description of a massive combinatorial space? Perhaps it's intractably large and normal brute force techniques won't work. Second, if that's true, do you have enough data to learn some sort of model of the topology of that space? Or maybe a simulator is available or learnable that can generate some additional synthetic data. Ideally, you have both. Third, you need a clear objective that you're trying to minimize or maximize. In games, that is winning or maximizing the score. In a natural system, that might be minimizing the free energy in that system. If you can quantify that, you can then use a model to search with the guidance of the objective function toward the optimal solution. It turns out a lot of problems fit that description.

Beyond that, we look for problems where the impact of solving it would be huge. If you think about the tree of all knowledge, there are some root node problems that, if you solve them, you unlock entire new branches of research. Protein folding was definitely one and it opens up opportunities in drug discovery and disease understanding. Whether fusion for energy, weather systems, material design, or fundamental mathematics – all of the problems we try to work on have those characteristics.[3]

Manyika. Say more about math. It is not intuitively obvious that all of mathematics has structure, at least not structure we understand. Distribution of prime numbers is perhaps a good example.

Hassabis. There's obviously still an open question about how much structure there is in numbers. I suspect it is quite a lot because math was developed to describe the real world, which does have structure, and it seems to be able to do that very well.

There are brilliant mathematicians like Srinivasa Ramanujan, who was incredibly intuitive and came up with amazing solutions without going through all the intermediate steps, so much so that he felt these leaps of imagination were divinely inspired. We call it implicit knowledge or intuition because we can't consciously access it or explicitly explain it in a structured way.

But intuition is not magic; it is just the word we use for the implicit knowledge the brain acquires through experiences. People who are very good at certain domains, say, like music, have extremely efficient and highly abstract versions of that knowledge. There's still a process there though, and I think in the limit it could be modeled by an AI system.

We're applying AI to mathematics and already seeing a glimpse of what can be modeled. Our recent Gemini model, Deep Think, reached gold medal–level performance in the International Mathematical Olympiad, which is a super difficult math challenge for the top high school students in the world.[4] The model is given a natural-language problem that it converts into a formal mathematical language, then solves it in an iterative way by performing some mathematical operations on the current working version of a proof until it reaches a verifiably correct proof.

But that's quite different from understanding if there are patterns in the numbers or structure in mathematics. That would require another kind of system. I think both are interesting – AI solving mathematical conjectures and finding deeper patterns in mathematics – and I think both may be possible.

Manyika. What gaps and limitations do you see in current AI systems when it comes to being able to do science? What new breakthroughs do you think are needed?

Hassabis. One thing that's missing and one of the reasons why I don't think today's systems are close to AGI (artificial general intelligence) yet is that they don't have true creativity. What I mean by that is, can you have an AI system that doesn't just solve a difficult problem, which is obviously quite useful and impressive, but can actually come up with a beautiful or meaningful conjecture like the Millennium Prize problems or the Riemann hypothesis? I think that's much more difficult.

Asking the right question is the hardest part of science. What's the right hypothesis that's worth exploring and worth doing science on? We don't fully understand how that level of creativity works in human experts, and we certainly don't have it yet in our AI systems.

The other missing feature in current AI systems is consistency. On the one hand, these systems can solve very complicated Math Olympiad problems; on the other hand, if you pose the question slightly differently, it can still mess up on elementary math. That wouldn't be the case for a human mathematician. They would have consistency of capability across the field.

Manyika. What does the idea of AGI mean to you?

Hassabis. I've always defined AGI as a system that exhibits all the cognitive capabilities the human mind has. That's a critical reference point because the human brain is the only existence proof that we know of so far, perhaps in the universe, that general intelligence is even possible. The brain is extremely general and adaptable – it's incredible that with minds evolved for hunter-gathering, we were able to invent the scientific method and build the marvels of modern civilization around us, from telescopes to jets to computers. The ingenuity of history's great artists, scientists, and philosophers all emerged from this same remarkable brain architecture. In my opinion, we won't really know for sure that we've built a fully general system until we're able to match the entirety of what we know that brain architecture can do.

From a theoretical point of view, this definition is closely related to the concept of a Turing machine, which Alan Turing famously conjectured could compute any algorithm given infinite time and memory. The brain can be viewed as a biological approximation to a Turing machine, meaning that, in theory, it could learn anything that is computable. However, as with any finite system, there always has to be some degree of specialization because there is only so much time, memory, and information. No practical system can be optimized for every possible task simultaneously; we can't circumvent the "no free lunch" theorem. The key, though, is the brain's potential to learn almost anything, especially as we are able to design and build tools and machines to help us gain knowledge.

Any candidate AGI system should have the consistency across the board to cover the incredible range of human abilities, including being capable of true intuitive leaps of discovery and coming up with a useful novel hypothesis or conjecture. This could be tested by, for example, training an AI system with a knowledge cutoff date of 1910 and seeing if it is able to come up with General Relativity like Einstein famously did in 1915. For now, the answer is clearly no.

What foundation models can already do today is extremely impressive, and in my opinion there's no doubt that they will ultimately form the backbone of any future AGI system. Whether or not scaling up these models further is enough on its own to reach AGI is an open empirical question, and I would not be surprised if a small number of breakthroughs on the level of transformers or deep reinforcement learning were still needed. So that's why at Google DeepMind we continue to relentlessly push forward both the frontiers of scaling and innovation in parallel.

Manyika. In assessing whether we've achieved AGI, on the one hand, one could use a battery of tests, and, on the other hand, perhaps there is some intuitive leap or spark that will happen – maybe like move thirty-seven (when AlphaGo went outside human strategy and intuition to play an incredibly unlikely move that led to victory over world champion Go player Lee Sedol) – and we'll say, "Ah, I think we've got it!" Is it both, or is it one or the other?

Hassabis. I think you need both. This is such an important moment in human history and I think we should be really scientific and sure about it. There should be quite a lot of obvious examples of invention or creativity happening, where some new physics is invented or some new conjecture is proposed that mathematicians agree is super meaningful. Instead of just coming up with a great new strategy in Go like AlphaGo did, could it invent a game as elegant as Go? I would want to see several examples of this happening for it to at least start even being a candidate.

Then I think we should have this exhaustive scientific approach to make sure there aren't any flaws or holes in the system. An example of that today is LLMs (large language models). It's very easy for the average person to come up with things they can't do. For example, they can't even play chess to an amateur level or learn a new game, and they don't continually learn. There are problems with reasoning, memory, and consistency. All of those would need to be solved in order to pass a test like the one I was proposing.

Manyika. How should scientists think about AI when it comes to discovery? Is AI a tool, a collaborator, something altogether different? And how do you imagine the role of scientists in the future when we have these more capable systems?

Hassabis. For the next decade or so, we should think about AI as this amazing tool to help scientists, whether that's number crunching, pattern matching, or some sort of low-level coding or engineering. Used in these ways, I think it is going to be the ultimate tool to help advance science and medicine and it will usher in a new golden era of discovery. Conjectures and hypotheses are going to come from the human experts, and then they'll have these tools that make them much more efficient – maybe ten times or one hundred times more – at what they do. Beyond that timeframe, it is hard to say with any certainty, but perhaps these systems will become more like collaborators.

Manyika. What does it mean to you to use AI in science responsibly? I know you care deeply about this, but what does that actually mean?

Hassabis. It means adhering to the scientific method and being extremely rigorous and careful with your tests and the claims made about the systems. It also means being very thoughtful about what you deploy and use the systems for and trying to think through their second-order effects. You're never going to get that perfect because it's such a new technology and the frontier is evolving so fast, but one should make every possible effort.

Related to that, it means using the scientific method itself to help with understanding AI: its limitations, its capabilities, and its risks. That understanding should then be used to put the right guardrails on AI, which are then monitored and studied as well.

I think the scientific method is the best approach to tackling the challenges with AI as it becomes more prevalent and more powerful, rather than perhaps a more hacker mentality of building first and asking questions later. That has its place, of course, and it's amazing for progress, but with something as transformative and as consequential as AI, we need to use the best tools we have, which in my opinion is the scientific method, to have as much foresight and understanding of these technologies as we can.

Manyika. What are the biggest benefits to humanity that you'd like to see from your work?

Hassabis. Number one is improving human health and curing diseases, especially those that disproportionately affect the poorer parts of the world. I think we can do so much better than we're doing today.

Then there is helping with energy, climate, and the environment by developing new materials and technologies faster, optimizing existing infrastructure and grids, and also modeling what's happening to the world and how we're affecting ecosystems. In my opinion, those are the most important things we can use AI for in the near term.

Manyika. If we're having this conversation in 2050, what are we going to be talking about?

Hassabis. Given the pace of how fast things are advancing, almost week by week, the only thing we know for certain is that by 2050 things will likely look very different. Hopefully the field will have safely stewarded AGI into the world for the benefit of all humanity, and we will be discussing how to navigate a post-AGI world of radical abundance and living in a new golden era of scientific discovery and wonder!

ABOUT THE AUTHORS

Demis Hassabis, a Member of the American Academy since 2021, is a Nobel laureate and an artificial intelligence researcher. He is Co-Founder and CEO of DeepMind and of Isomorphic Labs. In 2016, DeepMind's AlphaGo became the first program to beat a world champion at the complex game of Go. In 2020, its AlphaFold 2 program solved the fifty-year grand challenge of protein structure prediction.

James Manyika, a Member of the American Academy since 2019, is SVP at Google-Alphabet and President for Research, Labs, Tech & Society.

ENDNOTES

[1] David Silver, Aja Huang, Chris J. Maddison, et al., "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature* 529 (7587) (2016): 484–489; and John Jumper, Richard Evans, Alexander Pritzel, et al., "Highly Accurate Protein Structure Prediction with AlphaFold," *Nature* 596 (7873) (2021): 583–589.

[2] Josh Abramson, Jonas Adler, Jack Dunger, et al., "Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3," *Nature* 630 (8016) (2024): 493–500.

[3] Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, et al., "Probabilistic Weather Forecasting with Machine Learning," *Nature* 637 (8044) (2025): 84–90; and Yongji Wang and Sam Blackwell, "Discovering New Solutions to Century-Old Problems in Fluid Dynamics," Google DeepMind, September 18, 2025, https://deepmind.google/blog/discovering-new-solutions-to-century-old-problems-in-fluid-dynamics.

[4] Thomas Hubert, Rishi Mehta, Laurent Sartran, et al., "Olympiad-Level Formal Mathematical Reasoning with Reinforcement Learning," *Nature*, November 12, 2025 (unedited, prepublication manuscript), https://www.nature.com/articles/s41586-025-09833-y.