

Operations Practice

An executive primer on artificial general intelligence

While human-like artificial general intelligence may not be imminent, substantial advances may be possible in the coming years. Executives can prepare by recognizing the early signs of progress.

by Federico Berruti, Pieter Nel, and Rob Whiteman



Headlines sounding the alarms that artificial intelligence (AI) will lead humanity to a dystopian future seem to be everywhere. Prominent thought leaders, from Silicon Valley figures to legendary scientists, have warned that should AI evolve into artificial general intelligence (AGI)—AI that is as capable of learning intellectual tasks as humans are—civilization will be under serious threat.

Few seeing these warnings, stories, and images could be blamed for believing that the arrival of AGI is imminent. Little surprise, then, that so many media stories and business presentations about machine learning are accompanied by unsettling illustrations featuring humanoid robots.

Many of the most respected researchers and academics see things differently, however. They argue that we are decades away from realizing AGI, and some even predict that we won't see AGI in this century. With so much uncertainty, why should executives care about AGI today? The answer is that, while the timing of AGI is uncertain, the disruptive effects it could have on society cannot be understated.

Much has already been written about the likely impact of AI and the importance of carefully managing the transition to a more automated world. The purpose of this article is to provide an AGI primer to help executives understand the path to machines achieving human-level intelligence, indicators by which to measure progress, and actions the reader can take to begin preparations today.

How imminent is AGI?

In predicting that AGI won't arrive until the year 2300, Rodney Brooks, an MIT roboticist and co-founder of iRobot, doesn't mince words: "It is a fraught time understanding the true promise and dangers of AI. Most of what we read in the headlines... is, I believe, completely off the mark."

Brooks is far from being a lone voice of dissent. Leading AI researchers such as Geoffrey Hinton and Demis Hassabis have stated that AGI is nowhere close to reality. In responding to one of Brooks' posts, Yann LeCun, a professor at the Courant Institute of Mathematical Sciences at New York

University (NYU), is much more direct: "It's hard to explain to non-specialists that AGI is not a 'thing', and that most venues that have AGI in their name deal in highly speculative and theoretical issues..."

Still, many academics and researchers maintain that there is at least a chance that human-level artificial intelligence could be achieved in the next decade. Richard Sutton, professor of computer science of the University of Alberta, stated in a 2017 talk: "Understanding human-level AI will be a profound scientific achievement (and economic boon) and may well happen by 2030 (25% chance), or by 2040 (50% chance)—or never (10% chance)."

What should executives take away from this debate? Even a small probability of achieving AGI in the next decade justifies paying attention to developments in the field, given the potentially dramatic inflection point that AGI could bring about in society. As LeCun explains: "There is a thin domain of research that, while having ambitious goals of making progress towards human-level intelligence, is also sufficiently grounded in science and engineering methodologies to bring real progress in technology. That's the sweet spot."

For business leaders, it is critical to identify those researchers who operate in this sweet spot. In this executive's guide to AGI, we aim to help readers make that assessment by reviewing the history of the field (see sidebar, "A brief history of AI"), the problems that must be solved before researchers can claim they are close to developing human-level artificial intelligence, and what executives should do given these insights.

What capabilities would turn AI into AGI?

To understand the complexity of achieving true human-level intelligence, it is worthwhile to look at some the capabilities that true AGI will need to master.

Sensory perception. Whereas deep learning has

A brief history of AI

The term “artificial intelligence” was coined by John McCarthy in the research proposal for a 1956 workshop at Dartmouth that would kick off humanity’s efforts on this topic. The AI topics that McCarthy outlined in the introduction included how to get a computer to use human language; how to arrange “neuron nets” (which had been invented in 1943) so that they can form concepts; how a machine can improve itself (that is, learn or evolve); how a machine could form abstractions by using its sensors to observe the world; and how to make computers think creatively. In essence, McCarthy was describing in 1956 what we now call AGI.

McCarthy was certainly not the first to talk about machines and “intelligence.” Alan Turing’s 1950 paper on “Computing machinery and intelligence” introduced the “imitation game,” a test of a machine’s ability to exhibit intelligent behavior, and which became known as the “Turing test.” Turing optimistically estimated that, in the year 2000, a computer with 128Mb of memory would have a 70 percent chance of fooling a person. In his earlier 1948 paper on “Intelligent machinery,” he describes what we today call computers, as well as a machine that fully imitates a person. He points out that our ability to build adequate sensors and actuators might not be sufficient for some time and that our efforts are best invested in the aspect of intelligence that relates to games and cryptography. The clear aspiration, however, has always been to achieve human-level intelligence.

The early work, like Turing suggested, revolved around subject areas that do not require too much sensing and action, such as those of games and language translation. Research communities

around computer vision, natural language understanding, and neural nets are, in many cases, several decades old.

Within these AI research communities, there has been substantial success and advancement in what the field now terms “narrow AI” applications. Narrow AI is the application of AI techniques to a specific and well-defined problem, such as chatbots that help customers resolve issues with their phone bills, algorithms that spot fraud in credit-card transactions, and natural-language-processing engines that quickly process thousands of legal documents. Applying narrow AI solutions in use cases across industries can generate tremendous economic benefits. Our colleagues’ research shows that the potential value of applying this sort of deep learning could range from \$3.5 trillion to \$5.8 trillion annually.

To differentiate themselves from researchers solving narrow AI problems, a few research teams have claimed an almost proprietary interest in producing human-level intelligence (or more) under the name “artificial general intelligence.” Some have adopted the term “super-intelligence” to describe AGI systems that by themselves could rapidly design even more capable systems, with those systems further evolving to develop capabilities that far exceed any possessed by humans.

MIT roboticist Rodney Brooks describes the four previous attempts at AGI, along with their approximate start dates, in discussions that provide important context for understanding their progress.

Symbolic AI (1956). The key concept is the use of symbols and the encoding of

knowledge of the world through relationships between these symbols. One familiar example is the knowledge that a German shepherd is a dog, which is a mammal; all mammals are warm-blooded; therefore, a German shepherd should be warm-blooded.

The limitation is that humans still provide the ground truth by encoding our knowledge of the world, rather than allowing an AI system to observe and encode these relationships itself. Symbolic AI was the dominant paradigm of AI research from the mid-1950s until the late 1980s. One ongoing effort to provide a solution to common-sense reasoning through symbolic AI is the Cyc Project, launched in 1984 to collect knowledge represented declaratively in the form of logical assertions, of which it had collected 25 million by 2017.

Neural networks (1954, 1969, 1986, 2012). Recent advances in speech recognition (now widely used in devices such as smart speakers), and in computer vision (such as limited self-driving abilities in cars) are all due to deep neural networks. The artificial neuron as a computational model of the “nerve net” of the brain was proposed as early as 1943.¹ In the decades since then, it has been through multiple highs and lows in its popularity as a tool for AI.

MIT’s Marvin Minsky and Seymour Papert put a damper on this research in their 1969 book “Perceptrons,” where they mathematically demonstrated that neural networks could only perform very basic tasks. They also discussed the difficulty of training multi-layer networks. In 1986, however, Geoffrey Hinton and some colleagues solved this problem with the publication of the back-prop-

¹Warren McCulloch and Walter Pitts, “A logical calculus of the ideas immanent in nervous activity,” *Bulletin of Mathematical Biophysics*, volume 5, 1943.

agation algorithm. In the 1990's, work by Yann LeCun made major advances in neural networks' use in computer vision, and Jürgen Schmidhuber similarly advanced the application of recurrent neural networks as used in language processing.

In 2012, Hinton and two of his students highlighted the power of deep learning when they obtained significant results in the ImageNet competition. With the power of graphics processing units (GPUs), Hinton's back-prop algorithm could be applied to neural networks with substantially more layers, hence making them "deeper" and introducing the terms "deep neural networks" and "deep learning." This kicked off the past few years' focus on deep learning, almost to the exclusion of most other approaches.

Traditional robotics (1968). During the first few decades of AI, researchers also sought to build robots to advance research. Some robots were mobile, moving around on wheels, while others were fixed, with articulated arms. Robots used the earliest attempts at computer

vision to identify and navigate through their environments or to understand the geometry of objects and maneuver them, such as moving around blocks of various shapes and colors. Accordingly, although robots have been used in factories for decades, most rely on highly controlled environments with thoroughly scripted behaviors that they perform repeatedly. As valuable as that is, it has not contributed significantly to the advancement of AI itself.

One area where traditional robotics did have substantial impact is in the way robots build maps of their environment using only partial knowledge and observations, through a process called "simultaneous localization and mapping" (SLAM). SLAM algorithms became part of the basis for self-driving cars and are used in consumer products, from vacuum-cleaning robots to quadcopter drones. Today, this work has evolved into behavior-based robotics, commonly referred to as haptic technology, which provides a semblance of human touch, or physicality, to AI.

Behavior-based robotics (1985).

Insects navigate well in the real world with very few neurons. This observation inspired researchers to investigate how a robot could resolve conflicts when partial observation and knowledge of the physical world might provide conflicting commands to its actuators. A conflict-resolution mechanism needs to be heuristic in nature. As Brooks states, "Behavior-based systems work because the demands of physics on a body embedded in the world force the ultimate conflict resolution between behaviors, and the interactions."

The behavior-based approach starts out as reactive to how the world is changing. As a result, robots applying this approach are more robust in the real world, which is why the majority of deployed robots use it. Neural networks by themselves aren't capable of this level of adaptation. In each recent successful deployment, neural networks are embedded as part of symbolic AI systems or behavior-based systems.

enabled major advances in computer vision, AI systems are far away from developing human-like sensory-perception capabilities. For example, systems trained through deep learning still have poor color consistency: self-driving car systems have been fooled by small pieces of black tape or stickers on a red stop sign. For any human, the redness of the stop sign is still completely evident, but the deep learning-based system gets fooled into thinking the stop sign is something else. Current computer vision systems are also largely incapable of extracting depth and three-dimensional information from static images.

Humans can also determine the spatial characteristics of an environment from sound, even when listening to a monaural telephone channel. We can understand the background noise and form a mental picture of where someone is when speaking to them on the phone (on a sidewalk, with cars approaching in the background). AI systems are not yet able to replicate this distinctly human perception.

Fine motor skills. Any human can easily retrieve a set of keys from a pocket. Very few of us would let any of the robot manipulators or humanoid

hands we see do that task for us. Researchers in the field are working on this problem. A recent demonstration showed how reinforcement learning could teach a robot hand to solve a Rubik's cube. Although Claude Shannon built a robot to solve the cube decades ago, this demonstration illustrates the dexterity involved in programming robot fingers on a single hand to manipulate a complex object.

Natural language understanding. Humans record and transmit skills and knowledge through books, articles, blog posts, and, more recently, how-to videos. AI will need to be able to consume these sources of information with full comprehension. Humans write with an implicit assumption of the reader's general knowledge, and a vast amount of information is assumed and unsaid. If AI lacks this basis of common-sense knowledge, it will not be able to operate in the real world.

NYU professors Gary Marcus and Ernest Davis describe this requirement in more detail in their book "Rebooting AI," pointing out that this common-sense knowledge is important for even the most mundane tasks anyone would want AI systems to do. As Douglas Hofstadter notes, the fact that free machine-translation services have become fairly accurate through deep learning does not mean that AI is close to genuine reading comprehension, as it has no understanding of context over multiple sentences—something which even toddlers handle effortlessly. The various reports of AI passing entrance exams or doing well at eighth-grade science tests are a few examples of how a narrow AI solution can be easily confused for human-level intelligence.

Problem solving. In any general-purpose application, a robot (or an AI engine living in the cloud) will have to be able to diagnose problems, and then address them. A home robot would have to recognize that a light bulb is blown and either replace the bulb or notify a repair person. To carry out these tasks, the robot either needs some aspect of the common sense described above, or the ability to run simulations to determine possibilities, plausibility, and probabilities. Today, no known systems possess either such common sense, or such a general-purpose simulation capability.

Navigation. GPS, combined with capabilities such as simultaneous localization and mapping (SLAM), has made good progress in this field. Projecting actions through imagined physical spaces, however, is not far advanced when compared with the current capabilities of video games. Years of work are still required to make robust systems that can be used with no human priming. Current academic demonstrations have not come close to achieving this goal.

Creativity. Commenters fearing superintelligence theorize that once AI reaches human-level intelligence, it will rapidly improve itself through a bootstrapping process to reach levels of intelligence far exceeding those of any human. But in order to accomplish this self-improvement, AI systems will have to rewrite their own code. This level of introspection will require an AI system to understand the vast amounts of code that humans cobbled together, and identify novel methods for improving it. Machines have demonstrated the ability to draw pictures and compose music, but further advances are needed for human-level creativity.

Social and emotional engagement. For robots and AI to be successful in our world, humans must want to interact with them, and not fear them. The robot will need to understand humans, interpreting facial expressions or changes in tone that reveal underlying emotions. Certain limited applications are in use already, such as systems deployed in contact centers that can detect when customers sound angry or worried, and direct them to the right queue for help. But given humans' own difficulties interpreting emotions correctly, and the perception challenges discussed above, AI that is capable of empathy appears to be a distant prospect.

Four ways to measure progress

Instead of still trying to use the Turing Test, Brooks suggests four simple ways to measure our progress toward human-level intelligence that are summarized here below. Similarly, numerous companies and research organizations are exploring alternative frameworks to measure progress based on granular human-equivalent

capabilities, requirements to perform certain human tasks, or the combination of capabilities to perform every human job.

The object-recognition capabilities of a two-year-old

In the first case, two-year-old children who are only used to sitting on white chairs will realize that they can also sit on black chairs, three-legged brown stools, or even on rocks or stacks of books.

The language-understanding capabilities of a four-year-old

Four-year-olds are typically able to converse and follow context and meaning over multiple exchanges with a decent understanding as to the subtleties of language. We don't need to start every sentence by first stating their names (unlike today's "smart" speakers), and they can understand when a conversation has ended, or the participants have changed. Children can understand singing, shouting, and whispering, and perform each of these activities. They even understand lying and humor.

The manual dexterity of a six-year-old

Most six-year-olds are able to dress themselves and can likely even tie their own shoes. They can perform complex tasks requiring manual dexterity using a variety of different materials, and can handle animals and even younger siblings.

The social understanding of an eight-year-old

Eight-year-olds can hold their own beliefs, desires, and intentions, explaining them to others and understanding when others explain theirs. They can infer other people's desires and intents from their actions and understand why they have those desires and intents. We don't explain our desires and intents to children because we expect them to understand what they are observing.

Although the AI community is active in research to address all these aspects, we are likely decades away from achieving some of them. In more narrow applications, it seems plausible that object recognition, language understanding, and manual dexterity can be mastered to a sufficient extent in the medium term to address specific use cases.

Very often in the literature, the concept of a robotic elder-care robot is used as a conceptual test case. With the advances we're seeing, it's certainly plausible that a simplified and useful domestic robot that can offer some assistance to an elderly person might be available within the next decade, even if controlled by a remote human pilot at the beginning.

What advances could hasten inflection points?

The reduction in storage costs over the last two decades brought about the concept of "big data." The computing advances in GPUs uniquely enable an algorithm to be applied to much larger neural networks. With these neural networks trained on very large data sets, researchers accomplished all the recent advances brought about through deep learning. The combination of data, algorithms, and computing advances caused an inflection point. To look for the next AI inflection point, it is useful to consider the landscape again using those three component parts.

Major algorithmic advances and new robotics approaches. It may very well require completely new approaches to move us toward the level of intelligence displayed by a dog or a two-year-old human child. One example researchers are exploring is the concept of embodied cognition. Their hypothesis is that robots will need to learn from their environment through a multitude of senses, much like humans do in the early stages of life—and that they will have to experience the physical world through a body similar to that of humans in order to cognitively develop in the same way as humans do. With the physical world already designed around humans, there is merit in this approach. It prevents us from having to redesign so many of our physical interfaces—everything from doorknobs to staircases and elevator buttons. Certainly, as described in a previous section, if we are going to bond with smart robots, we are going to have to like them. And it is likely that such bonding is only going to happen if they look like us.

The entire advance in deep learning is enabled by the backpropagation algorithm, which allows

large and complex neural networks to learn from training data. Hinton, along with colleagues David Rumelhart and Ronald Williams, published “Learning representations by back-propagating errors” in 1986. It took another 26 years before an increase in computing power and the growth in “big data” enabled the use of that discovery at the scale seen today. Whereas a multitude of researchers have made improvements in the way backpropagation is used in deep learning, none of these improvements has been transformative in the same way. (Hinton’s more recent work on “capsule networks” may very well be one such algorithmic advance which could, among other applications, overcome the limitations of today’s neural networks in machine vision.)

Deep learning assumes a “blank slate” state, and that all “intelligence” can be learned from training data. Anyone who has ever observed a mammal being born would recognize that something like a fawn starts life with a level of built-in knowledge. It stands within 10 minutes, knows how to feed almost immediately, and walks within hours. As Marcus and Davis point out in *Rebooting AI*, “The real advance in AI, we believe, will start with an understanding of what kinds of knowledge and representations should be built in prior to learning, in order to bootstrap the rest.” The recent success of deep learning may have drawn away research attention from the more fundamental cognitive work required to make progress in AGI.

Major computing advancements. The application of GPUs to training deep neural networks was a critical step-change that made the major advances of the last several years possible. GPUs uniquely enabled the complex calculations required by Hinton’s backpropagation algorithm to be applied in parallel, thereby making it possible to train hugely complex neural nets within a finite time. Before any further exponential growth toward AGI can be expected, a similar inflection point in computing infrastructure would need to be matched with unique algorithmic advances.

Quantum computing is often touted as one of the potential computing advances that could change our society. But, as our colleagues recently noted in a research report, quantum computing is

proposed not as a replacement for today’s devices, but for highly complex statistical problems that current computing power cannot address. Moreover, the first real proof that quantum computers can handle these types of problems occurred only in late 2019, and only for a purely mathematical exercise with no real-world use at all. The hardware and software to handle problems such as those required for advancements in AI may not arrive until 2035 or later. Nonetheless, quantum computing remains one of the most likely possible inflection points and one to keep close tabs on.

Substantial growth in data volume, and from new sources. The rollout of 5G mobile infrastructure is one of the technology advances touted to bring about a significant increase in data due to the way the technology can enable growth in the internet of things (IoT). Research conducted by our colleagues has, nevertheless, noted roadblocks to 5G implementation, particularly in the economics for operators. Also, in a 2019 survey, operators reported that they did not see IoT as a core objective for 5G, because the existing IoT capability was likely sufficient for most use cases. As a result, 5G appears unlikely by itself to serve as a major inflection point for increasing data volume and as a subsequent enabler of training data. Most of the benefits may already have appeared.

New robotics approaches can yield new sources of training data. By placing human-like robots with even basic functions among humans—and doing so at scale—large sets of data that mimics our own senses can help close a training feedback loop that enhances the state of the art. Advanced self-driving cars are one such example: the data collected by cars already on the market are acting as a training set for future self-driving capability. Furthermore, much research is being done in human-robot interaction. By finding initial use cases for human-like robots, this research could greatly add to the training data necessary to expand their capabilities.

What executives could do

What are the next steps for executives? The best way to counteract the hype about AGI is to take tangible actions to monitor developments and position your organization to respond appropriately

to real progress in the field. The following checklist offers categories of actions to consider.

- **Stay closely informed** about developments in AGI, especially with regard to the ways AGI could be advancing more rapidly than expected. To enable this, connect with start-ups and develop a framework for rating and tracking progress of AGI developments that are relevant to your business. Additionally, begin to consider the right governance, conditions, and boundaries for success within your business and communities.
- **Tailor environments** to enable narrow-AI advances now—don't wait for AGI to develop before acting. A number of steps can be taken today to adjust the landscape and increase uptake. These include simplifying processes, structuring physical spaces, and converting analog systems and unstructured data into digital systems and structured data. The digital and automation programs of today can smooth the transition to AGI for your customers, employees, and stakeholders.
- **Invest in combined human-machine interfaces** or “human in the loop” technologies that augment human intelligence rather than replace it. This category includes everything from analytics to improve human decision making to cognitive agents that work alongside call-center agents. Using technology to help people be more productive has been the engine of economic progress and will likely remain so for the foreseeable future.
- **Democratize technology** at your company, so progress is not bottlenecked by the capacity of your IT organization. This does not mean letting technology run wild. It means building technical capabilities outside of IT, selectively deploying platforms that require little or no coding skills, and designing governance models that encourage rather than stifle innovation.
- **Organize your workers** for new economies of scale and skill. The rigid organization structures and operating models of the past are poorly suited for a world where AI is advancing rapidly. Embrace the power of humans to work in complex environments and self-organize. For example, institute flow-to-the-work models that allow people to move seamlessly between initiatives and groups.
- **Place small bets** to preserve strategic options in areas of your business that are most exposed to AGI developments. For example, consider investments in technology firms pursuing ambitious AI research and development projects in your industry. It's impossible to know when (or if) your bets will pay off, but targeted investments today can help you hedge existential risks your business might face in the future.
- **Explore open innovation models** and platforming with other companies, governments, and academic institutions. Such arrangements are essential to test the art of the possible and the business nuances of AGI development. It's hard to keep up with the rapidly changing AGI landscape without firsthand experience working alongside leading organizations.

AGI may not be ready this decade or even this century—but some of the capabilities may start appearing in places you might not expect. The benefits will accrue most to those who are observant—and prepared.

Federico Berruti is a partner in McKinsey's Toronto office, **Pieter Nel** is an alumnus of the New York office, and **Rob Whiteman** is a partner in the Chicago office.