Received: May 22, 2025
Tracking No. maz-gwfw-m9ja
Comments Due: May 28, 2025 **Submission Type:** Web

PUBLIC SUBMISSION

Docket: NSF-2025-OGC-0001
NITRD_FRDOC_0001

Comment On: NSF-2025-OGC-0001-0001
Request for Information: Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

Document: NSF-2025-OGC-0001-DRAFT-0095
Comment on FR Doc # 2025-07332

Submitter Information

Name: William Marcellino

General Comment

As a research scientist actively involved in the development of AI systems and tools, as well as researching public policy on the development and deployment/diffusion of AI systems, I'm excited by the prospect of a coherent US strategy for global AI leadership. AI, and the possibility of artificial general intelligence (AGI) is such a potentially transformative technology that the US cannot afford to cede leadership in this area to other actors geopolitically, especially China, who are aggressively pursuing a coherent strategy towards achieving AI dominance and AGI.

Currently, the lack of a US AI strategy has left AI development in the hands of proprietary US frontier labs, and while this has led to many advances, has also meant a brittle and potentially misaligned development pathway for US AI dominance. One problem is that the labs are hyper focused on the transformer architecture/LLM's as a pathway to AGI. However, there are inherent, potentially insurmountable barriers to scaling up transformers to AGI. As the attached document shows, the wider US research community understands that multiple complementary AI technologies will be required to achieve AGI. Secondly, the US is defaulting to a proprietary model paradigm, while China pursues global leadership in the open-source world. At least currently, open source and open weight AI has almost entirely closed the gap with the proprietary models of US labs. Making these open source/weight models attractive is part of a coordinated Chinese strategy to create an alternative AI ecosystem, with an alternative technology and hardware stack to compete with the US. Finally, leaving US AI development in the hands of private actors means that while development and allocation of capital is likely to be efficient from a commercial standpoint, it may not align in all ways with US strategic priorities. US government will likely have geopolitical and national security needs that are distinct. Just as in the Apollo space program, where the US was able to coordinate and synchronize the power of US industry to achieve a national goal, we may need something similar for sustained AI leadership.

As a result of the above factors, the US has made a potentially brittle bet on the future: it only works if US proprietary labs can somehow defy the odds and scale up LLM's to AGI. Every other future outcome looks like Chinese strategic dominance. The US government should hedge our national bets with a diverse portfolio of AGI/AI development approaches that supports both proprietary AI development, but also competes with China for open-source AI leadership globally. So, for example, the US government could use funding such as compute credits to support diverse lines of research across US universities. US universities are just as capable of innovation and competition as Chinese universities; however, our universities are relatively underfunded and under-resourced. And because China has a lead in critical AI areas, such as computer vision, robotics/embodied intelligence, brain inspired neuromorphic computing, and cognitive intelligence, we have a compelling national interest in achieving US leadership in these critical areas. Additionally, the US can pursue a more robust AI leadership role by supporting open source AI development. We want the rest of the world coming to our infrastructure and software stack to do AI innovation, development, and commercially valuable deployment.

The US can achieve and maintain global AI dominance, but only If we invest in a broad portfolio of AI technologies, that covers a multitude of uncertain possible futures.

Attachments

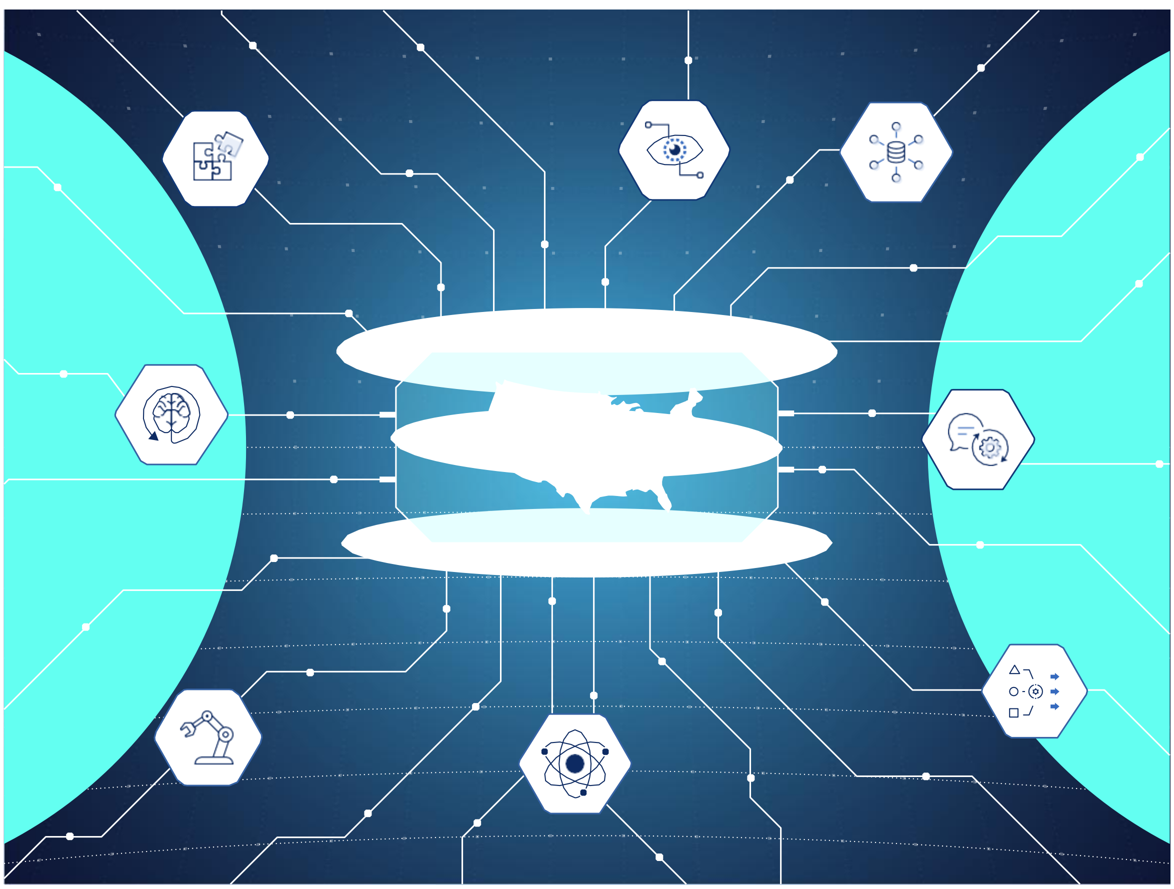
Charting Multiple Courses to Artificial General Intelligence

RAND

Expert Insights
PERSPECTIVE ON A TIMELY POLICY ISSUE

WILLIAM MARCELLINO, LAV VARSHNEY, ANTON SHENK, NICOLAS M. ROBLES, BENJAMIN BOUDREAUX

Charting Multiple Courses to Artificial General Intelligence



April 2025

For more information on this publication, visit **www.rand.org/t/PEA3691-1**.

Published by the RAND Corporation, Santa Monica, Calif.

Cover: Composite design by Carol Ponce adapted from images by Azat Valeev/Adobe Stock, MooseD/Adobe Stock, Smile3377/Adobe Stock.

About This Paper

Artificial intelligence (AI) and the potential emergence of artificial general intelligence (AGI) have important national security implications for the United States, particularly with regard to its competition with China. Existing AI technology—in the form of large language models (LLMs)—has shown great promise, and many in the AI technology and policy worlds think that LLMs may scale up to AGI in the near future. This paper is intended to complicate that position, explaining why there are barriers to LLMs hyperscaling to AGI, and why AGI may instead emerge from a suite of complementary, if not alternative, algorithmic and computing technologies. Our goal is to provide U.S. policymakers with a clear, nontechnical introduction to the issue of LLM hyperscaling and alternative pathways to AGI. We argue that there may be multiple courses to AGI and thus recommend that policy around AI avoid over-optimizing for a given possible future (e.g., hyperscaling LLMs), even while that policy addresses the possible near-term emergence of AGI in the hyperscaling paradigm.

Technology and Security Policy Center

RAND Global and Emerging Risks is a division of RAND that delivers rigorous and objective public policy research on the most consequential challenges to civilization and global security. This work was undertaken by the division's Technology and Security Policy Center, which explores how high-consequence, dual-use technologies change the global competition and threat environment, then develops policy and technology options to advance the security of the United States, its allies and partners, and the world. For more information, contact .

Funding

This effort was independently initiated and conducted within the Technology and Security Policy Center using income from operations and gifts from philanthropic supporters, which have been made or recommended by DALHAP Investments Ltd., Effektiv Spenden, Ergo Impact, Founders Pledge, Charlottes och Fredriks Stiftelse, Good Ventures, Jaan Tallinn, Longview, Open Philanthropy, and Waking Up Foundation. A complete list of donors and funders is available at www.rand.org/TASP. RAND donors and grantors have no influence over research findings or recommendations.

Acknowledgments

We would like to thank Joel Predd of RAND for his guidance and help making this a more-focused and clear paper. We also want to thank Andrew Lohn of the Center for Security and Emerging Technology and Timothy Gulden of RAND for their technical review and guidance.

Contents

Charting Multiple Courses to Artificial General Intelligence

Getting to Artificial General Intelligence from Large Language Models

The large language model (LLM) is a transformative technology that has generated an enormous amount of attention and investment dollars as companies across the globe race to build ever-larger, more-performant models trained on datasets of ever-increasing size. The introduction of ChatGPT in 2022 was an electrifying moment, and the prospect of artificial intelligence (AI) that uses natural language to solve problems and interact with humans—along with complementary trends in big data and compute—have begun a new *AI summer* with profound social and economic effects (Vöpel, 2024). This rapid progress in AI has inspired historic investment in AI, with U.S. business capital allocations topping $1 trillion in generative AI, in anticipation of increasing performance of LLMs (Goldman Sachs, 2024).

In this paper, we define *artificial general intelligence (AGI)* in the following three ways:

- *human or superhuman capabilities* across a variety of cognitive and metacognitive tasks
- deployed AI systems that perform *economically valuable* work by substituting automation for labor
- AI systems that display *emergent properties*, such as learning new skills and conducting new tasks.

The monetary stakes for scaling and improving these models is also high; existing generation models, such as GPT-4, cost upward of $78 million to train (Buchholz, 2024), and there is speculation that larger models could cost in the billions of dollars (Nesov, 2024). This massive investment reflects the potential economic and scientific impact of LLMs: They are likely to dramatically boost the productivity of workers across a variety of industries (Eloundou et al., 2023; Korinek, 2024) and may also revolutionize the pace of scientific discovery (Ifargan et al., 2024).

Many in industry and AI research argue that the trend toward larger and larger LLMs with increasing performance on a variety of benchmark tests means that the age of AGI is coming.[1] This is the *hyperscaling paradigm*: LLM systems will continue to grow in size and performance until they are self-improving, creating an irreversible launch to superhuman AGI (Aschenbrenner, 2024). Given the

[1] For more on AGI as defined by *capabilities*, see Morris et al., 2024. For more on AGI as defined by *deployed systems that perform economically valuable work*, see OpenAI, undated. For more on AGI as defined by *emergent properties*, see Chollet, 2019.

rapid pace of development and improvement in LLMs, if AGI is imminent (say, in the next two to four years, as some forecast), it will likely come from hyperscaling LLMs (Aschenbrenner, 2024). To some AI experts, this prospect entails manageable risks and an intelligence explosion leading to a post-scarcity world (Drexler, 2019), while others fear that this poses a genuine existential risk to humanity (Kruppa and Seetharaman, 2024).

In addition to the possibility of enormous economic gains or grave threats, the prospect of AGI has huge geopolitical stakes, particularly in terms of the U.S. competition with China. Because AGI has the potential to stimulate transformative economic growth and national power, there is a real concern that whoever gets to AGI first may have a permanent, decisive geopolitical advantage (Gill, 2020). The Chinese Communist Party, the governing party in China, has made a public commitment to global AI leadership, launching an ambitious, whole-of-government effort to develop AI through public-private partnerships that bring together national government, regional governments, academic researchers, and industry to accelerate AI development across a variety of AI technologies (Zhang and Luo, 2024). Clearly, the Chinese Communist Party takes competition for AI dominance seriously, and there is the potential for a race to AGI. U.S. government policy could have a huge effect on whether the United States maintains or extends its global AI lead.

Given these high stakes, it is important to understand the plausibility of the hyperscaling paradigm. If the hyperscaling paradigm is not plausible, then U.S. government policy on AI assumes that hyperscaling may be brittle, if not a source of strategic regret. In the rest of this paper, we lay out how existing AI research supports the idea that using LLMs may be an *important but insufficient* technical path to AGI. LLMs appear to have inherent limits that may require complementary technology to continue toward AGI, and while we cannot, at this time, make a confident claim about how to get to AGI, we do assert that the path to AGI is uncertain. U.S. government policymakers can consider the possibility that AGI emerges soon in the hyperscaling paradigm. But because of this uncertainty, the U.S. government may not want to put all its chips on black and bet the house, and instead, could prudently hedge against an uncertainty with policy that supports multiple possible futures.

Large Language Models Keep Getting Better, but Is That Enough?

Existing research suggests that hyperscaling may not be a viable or the only path to AGI. For example, while leading AI labs marketing their LLM products point to improving performance on reasoning benchmarks as LLMs scale, recent research shows that although this is true, scaling also leads to increasingly confident, wrong answers (Zhou et al., 2024). Whether these LLMs have any capacity to truly understand language—as opposed to *shortcut learning* of surface patterns—is broadly contested in the AI research world (Mitchell and Krakauer, 2023). Prominent AI researchers, such as François Chollet (2024) and Melanie Mitchell (2024), argue that LLMs are not meaningfully intelligent and fail when presented with data outside what they have memorized: They cannot do the *general* part of AGI. So, while there has been continuing improvement in LLM technology, the technology's development remains a complicated story, as we detail in the following paragraphs.

At the heart of the hyperscaling paradigm is the idea of *emergent abilities*. Early research on LLMs found unexpected and profound jumps in performance: Somehow, as models scaled, new abilities emerged without specific training for those abilities (Woodside, 2024). Subsequent research has shown, however, that emergence may be a mirage caused by faulty metrics. Benchmarks showing emergence were all-or-nothing measures and so steady, partial improvement toward problem-solving hid smooth improvement. When metrics are adjusted to measure progress and partial solving, improvements smooth out, and new abilities vanish (Schaeffer, Miranda, and Koyejo, 2024). Still others argue there may be true emergence with scaling but that LLMs may plateau in capabilities (Nayak and Varshney, 2024).

Furthermore, for usage of LLMs to be a viable path to AGI, LLMs likely need to go beyond statistical modeling of language to having the ability to reason in logic and math. Although LLMs have improved on problem-solving reasoning benchmarks as they scale, this may be the result of pattern memorization. One example of this is the *reversal curse*, in which models can memorize a relationship unidirectionally but not bidirectionally (Berglund et al., 2023; Golovneva et al., 2024). That is, LLMs can memorize that "A has feature B," but not that "B is a feature of A," unless the model is double-trained to separately memorize this relationship. Recent research on mathematical reasoning also highlights the issue of LLM performance as memorization (Mirzadeh et al., 2024). If benchmarks are abstracted to symbols (e.g., instead of "if Tony has four apples and Janet has six," the question has "if {name} has {x} apples and {name} has {y}") not only does accuracy drop dramatically (up to 65 percent), but this fragility also increases with the length of the benchmark question. Furthermore, if linguistically similar but irrelevant information emerges ("five of the kiwis are *smaller* than average"), LLMs tend to naively incorporate this irrelevant information—for example, by subtracting the smaller kiwis.

In addition to performance limits, there are potential economic limits to scaling, in particular data and energy constraints. At existing model scaling rates, the entire stock of human-generated training data may be exhausted within the decade, and ever-increasing quantities of AI-generated content may contaminate future training to the point that models collapse into lower and lower quality and diversity of output. This *data wall* may present serious challenges to scaling LLMs (Aschenbrenner, 2024; Villalobos et al., 2024). Beyond data constraints, the ever-growing power demands for training models and running those models (*inferencing*) may mean that researchers hit an *energy wall* that limits LLM scaling (Kurshan, 2024; Stojkovic et al., 2024). It is possible that data and energy walls could be surmounted through technical means,[2] but there are still credible reasons that hyperscaling may not work.

In summary, while LLM performance on benchmarks increases as LLMs scale, recent research raises the possibility this may be a function of higher-power memorization of patterns. Scaling and training result in increasingly confident—yet incorrect—answers to hard questions for multiple families of LLMs. Emergent abilities in LLMs may be artifacts of flawed measurement. Additionally, LLMs struggle with formal reasoning and math, apparently relying on rote memorization when faced with abstract or linguistically complex problems. Finally, from a practical standpoint, data and energy walls may economically constrain LLM scaling. Of course, none of this is definitive: It is still possible that hyperscaling will lead to AGI, and if AGI emerges in the near future, it is likely to come from

[2] The recent attention to DeepSeek's low-cost models is a good example of how algorithmic innovation can reduce energy demands. DeepSeek engineers were able to dramatically improve training efficiency and thus reduce compute (and, thus, energy) requirements for training their DeepSeek-V3 model (Liu et al., 2024).

scaling up LLMs. However, existing research raises important questions about the limitations of hyperscaling as a path to AGI. AGI might depend on additional, alternative technology.

What Else Might We Need for Artificial General Intelligence?

If hyperscaling LLMs is not a viable path to AGI, what might be? In this section, we survey a variety of complementary alternative technologies that may fill in the gaps that LLMs leave. Table 1 lists and describes a sample of some of the more promising complementary AI technologies that might fill in the gaps from LLMs. This list is not exhaustive but rather illustrative of how different AI technologies might contribute to AGI. This list is derived from a workshop that RAND researchers conducted with leading AI experts across U.S. federally funded research and development centers, university-affiliated research centers, and academia.

Table 1. Potentially Instrumental Artificial General Intelligence Technologies

Approach	Description
Physics or causal hybrids	• Integrates physical laws and causal reasoning into AI models to improve the models' utility in the real world.
Cognitive AI	• Mimics the human brain's architecture to improve efficiency and processing speed.
Information lattice learning	• Creates interpretable representations of patterns. • Facilitates robust learning from small datasets for novel, unseen problems.
Reinforcement learning	• Trains models through trial and error to learn optimal behaviors (e.g., in robotics, gaming, and autonomous systems) to train agents.
Neurosymbolic architectures	• Combines neural networks with symbolic reasoning to enhance interpretability and logic and mathematical problem-solving.
Embodiment	• Learns spatial relationships, object dynamics, and physical relationships through interaction (e.g., learning robots with sensors).
Neuromorphic computing	• Uses spiking neural networks for energy-efficient computation.

Potentially Instrumental Algorithmic and Computing Technologies

In the following sections, we unpack algorithmic and computing technologies, explaining them at the conceptual level and explaining how they might address fundamental limitations in existing AI systems and how they might contribute to achieving AGI.

Physics-Informed Neural Networks and Causal Models

Picture a child running through a home clutching barber's shears in their hand. This is likely a very uncomfortable image because adult humans have developed robust models for real-world physics and causality. That is, we have a good understanding of what happens when sharp steel hits soft flesh with force. We also have a good understanding of the disastrous consequences of such an event. Part of what allows people to be *generally intelligent* is their understanding of how the physical world works.

Physics-informed neural networks (PINNs) model the fundamental laws of physics, such as Newton's laws of motion. PINNs can solve problems, such as dynamics (e.g., simulating turbulence or predicting structural stress) and are highly valuable when data are sparse by leveraging prior knowledge of physics principles (Cuomo et al., 2022; Raissi, Perdikaris, and Karniadakis, 2019). *Causal models*, on the other hand, are for understanding cause-and-effect relationships. Causal models can distinguish between correlation and causation, and they allow AI systems to simulate counterfactual scenarios—for example, "What would happen if the car brakes were applied more forcefully?" Causal models are critical for safety because autonomous systems (e.g., robots, cars) navigate dynamic, real-world environments (Kacianka et al., 2019) and for the safety of potential AGI systems (Everitt et al., 2019; Holtman, 2021).

Although LLMs capture patterns in textual or visual data, there is no explicit representation of real-world physics or causality. Existing LLMs model "sharp objects" as a probabilistic relationship between words without any apparent grounding in the physical world. An LLM may model that "knife" lives near "cut" and "blood," but as of this writing, no LLM appears to simulate the dynamics of a knife's cutting or predict the potentially deadly consequences of a cut in physical terms. PINNs and causal models, however, bridge this gap in real-world understanding. While existing language models offer only surface-level semantic understanding (Vafa et al., 2024), PINNs can model physical processes, and causal models can reason about the outcomes of those processes. Together, PINNs and causal models could enable more-robust decisionmaking: an AI system that, for example, understands the danger of moving massively heavy pallets around frail human bodies in a warehouse.

Cognitive Artificial Intelligence

Cognitive AI takes human cognition as the starting point for AGI. In contrast with statistical and generative approaches, cognitive AI is designed to artificially reproduce the hallmark features of human intelligence. The goal of cognitive AI research is to engineer systems that are similar to human intelligence; specifically, systems that can learn concepts by interacting with the environment and other actors, have short- and long-term memory, can adaptatively learn how to act in different contexts, and can learn continuously and iteratively (Voss and Jovanovic, 2023).

While different research threads use terms of art, such as cognitive computing, cognitive AI, and artificial cognition, they all have in common a focus on human-like *cognition*. On the one hand, AI does not necessarily require cognition; for example, reinforcement learning is, in one sense, a kind of brute-force path to learning optimization, and LLMs are statistical models of patterns in data, such as language or proteins. On the other hand, cognitive AI requires a human-like capacity for thinking as a process to solve real-world problems by making sense of data in context (Sandini, Sciutti, and Morasso, 2024). This approach is designed to achieve a more-holistic form of AI in which machines can engage in reasoning, problem-solving, and decisionmaking in a way that mirrors human cognitive abilities (Sreedevi et al., 2022).

This AI technology could address the fundamental challenge of replicating the broad and adaptable intelligence that humans possess. One definition of the *G* in AGI includes the assumption that such intelligence has the ability to understand and learn from diverse experiences, make context-aware decisions, and apply knowledge flexibly across different domains—capabilities that are inherent to human cognition. By focusing on replicating these cognitive processes, cognitive AI provides a pathway to developing systems that can not only perform specific tasks but also generalize such learning to new, unforeseen challenges. This adaptability and contextual understanding may be essential for AGI to operate effectively in the complex, dynamic environments that characterize the real world.

Information Lattice Learning

If we threw a set of irregularly sized, differently colored, square- and star-shaped blocks on the ground, even a very young child could immediately detect the pattern that distinguishes the two classes. Furthermore, a child of a certain age could articulate a rule for classifying the blocks: Squares have four corners and equal sides; stars have more than four corners and can have different-length sides. LLMs lack this human-like ability to recognize patterns from a single or very few examples and then explain the rules of the pattern in human-understandable ways. However, a novel kind of AI, *information lattice learning*, does exactly this without using neural networks. This form of AI can discover known laws of music theory, chemistry, genetics, and quantum physics from very small amounts of data in the same human-interpretable form as textbooks, and it can also make new discoveries beyond what scientists have previously considered without human engineering to explicitly input any domain knowledge in advance (Yu, Evans, and Varshney, 2023).

Such general knowledge discovery could be used downstream for diverse applications: state-of-the-art classification of visual targets, semantic compression for 6G wireless, or helping people create ideas and artifacts that have never been imagined before. Because essentially no domain knowledge is needed in advance and few data are needed for training, information lattice learning captures a key aspect of general intelligence (Chollet, 2019). In fast-changing or idiosyncratic settings that arise in intelligence, defense, biosecurity, and other domains of national competitiveness, including research and development, the data efficiency and human controllability of information lattice learning have capabilities that are powerfully complementary to LLMs.

Reinforcement Learning

Reinforcement learning (RL) trains AI through trial and error rather than by learning rules or theory. So, for example, existing AI chess systems that beat any human player do not learn chess in a human way through a set of principles and strategies but rather by trying potentially millions of combinations to find optimal solutions. When an RL model makes progress, its policy is reinforced through a reward while suboptimal moves are penalized until a system, such as a chess AI expert, might see dozens of moves ahead down obscure paths that result in a small win, such as a single pawn taken. While this is a very nonhuman approach, it can be a powerful and useful one for specific tasks and within certain domains.

A paradigm example of this is an autonomous vehicle, such as a drone. With sufficient RL, drones (and other agents) can learn how to navigate complex terrain and dynamic situations without any human supervision; for example, aerial delivery drones can safely avoid power lines, and reconnaissance drones can hug terrain to avoid detection (AlMahamid and Grolinger, 2022). Because RL involves so many trial-and-error attempts, much RL training is done virtually (for example, by crashing a drone virtually over and over again but in a time-accelerated way to get enough experience to learn optimal policies). Beyond autonomous movement, RL has value in diverse areas, such as medical diagnosis and education (Radmehr, Singla, and Käser, 2024; Yu et al., 2023).

RL may be critical for developing AGI because it gives machines the ability to learn and adapt through experience, much like humans do. RL systems can tackle a wide variety of tasks and environments by continuously refining their actions based on feedback. Furthermore, RL could be combined with LLMs to create hybrid systems that reflect deep learning on a task (RL) with the ability to problem-solve (LLMs) (Pternea et al., 2024; Radmehr, Singla, and Käser, 2024), and, in fact, the most-recent tranche of reasoning LLMs from such companies as OpenAI and DeepSeek incorporate RL for such tasks as math, coding, and science question-answering (Mercer, Spillard, and Martin, 2025). RL's adaptability and continuous learning are key components in creating machines that could learn to act across diverse situations and tasks, moving humanity closer to achieving AGI.

Neurosymbolic Architectures

Neurosymbolic architecture refers to an emerging field that integrates the strengths of neural networks with symbolic reasoning, aiming to overcome the limitations of purely data-driven models, such as LLMs. An LLM is an example of a traditional neural network, which excels in pattern recognition and data-driven tasks but struggles with tasks that require abstract reasoning, logical inference, and generalization beyond training data, especially in advanced mathematics. Symbolic AI, on the other hand, uses formal logic and explicit knowledge representations (e.g., rules, ontologies) to reason about the world. By combining these two approaches, developers can design a hybrid neurosymbolic AI system that leverages the flexibility of neural networks in processing raw data with the interpretability and structured reasoning of symbolic systems (Garcez, Lamb, and Gabbay, 2019). This integration allows more-robust problem-solving across a wider variety of domains, including those that require common-sense knowledge and complex reasoning.

In contrast with LLMs, which rely purely on statistical learning from vast corpora of text data, neurosymbolic AI combines data-driven learning with explicit representations of knowledge. While LLMs have demonstrated impressive language capabilities, they remain limited by their reliance on pattern-matching rather than logical reasoning or understanding of the world (Bender et al., 2021). Neurosymbolic AI, by incorporating symbolic components, such as logical reasoning and structured knowledge, enables models to better handle such tasks as deductive reasoning, problem decomposition, and explanation generation (Zhang and Sheng, 2024). These models bridge the gap between the data-driven strength of neural networks and the structured intelligence of symbolic reasoning, allowing more-generalizable and explainable AI systems.

Neurosymbolic AI may be a critical step toward achieving AGI because of its ability to combine flexible learning with structured reasoning. AGI requires not just vast data processing capabilities but also the ability to reason, learn from fewer examples, and generalize knowledge across different domains—qualities that are difficult to achieve with purely neural-based models, such as LLMs (Schmidhuber, 2022). The ability to integrate symbolic reasoning into neural networks equips the system with higher-order cognitive capabilities, such as understanding context, forming causal relationships, and applying learned knowledge to new situations. This hybrid approach allows more human-like flexibility in thinking, such as dealing with incomplete information, explaining decisions, and reasoning about novel situations.

Embodiment

What if an AI system could learn about the world through interaction, similar to how a baby learns? Imagine a robot with a variety of sensors: cameras, microphones, hands with tactile sensors, maybe even chemical sensors for taste and smell. The robot roams around, learning about *dogs* by hearing them bark, feeling their fur, seeing them move, and smelling wet dog. The robot interacts with people (and maybe other AI agents) to learn through embodied interactions.

Embodied AI systems are grounded in reality, developing a nuanced understanding of spatial relationships, object dynamics, and physical interactions (Duan et al., 2022). This allows these systems to learn through interactions, much as humans do, and like humans, gain context-aware knowledge, which is something that LLMs lack. This high-context awareness is critical for AI systems to be able to make real-time decisions and adapt dynamically. And by bridging the gap between sensing and moving or manipulating (Hughes et al., 2022), embodied AI systems could have the physics, causal, and interactional knowledge to make sense of the consequences of their actions in the physical world and thus make more-informed and safe decisions.

Embodiment brings AI systems closer to human-like intelligence by enabling them to experience the world in a way that is similar to how humans do. This experiential aspect is valuable for developing empathy, intuition, and other cognitive abilities that are difficult to achieve through data processing alone.

Neuromorphic Computing

LLMs use incredible amounts of energy, not just for training large models but also when deployed and generating responses (*inferencing*), especially with new chain-of-thought techniques for reasoning using RL: *test-time* scaling rather than *train-time* scaling (Mercer, Spillard, and Martin, 2025). Because LLMs are a kind of artificial neural network, they require high-performance computer chips, specifically graphics processing units (GPUs). These GPUs can efficiently handle the complex matrix math of LLMs, but they have high energy demands to train and run. These power demands are a particular sustainability concern as LLMs hyperscale. Additionally, there is a single dominant supplier (the Nvidia Corporation) for GPUs, which raises cost and supply chain risks (Tang and Zhu, 2024).

While these traditional chips use clock-driven timing and constant power, a new class of neuromorphic chips has been developed that uses discrete (and thus low levels of power) electrical pulses for computation. Neuromorphic chips take inspiration from the human brain and use spiking signals and massive parallel connectivity among computational units to be vastly more energy efficient than traditional computer circuits. These chips have been deployed in labs but are not in widespread use as of this writing (early 2025). A different, nascent approach would be to culture actual biological neurons into artificial computing devices, so that such devices run on small amounts of sugar rather than large amounts of electricity (Zhang et al., 2024). If pushed to an industrial scale, these alternate physical substrates may upend supply chains for AI compute and change the energy calculus for AI.

A Robust Strategy That Covers Multiple Futures

In the previous section, we introduced a variety of possibly fruitful alternative technical paths to AGI. Our point is not to make specific recommendations but to provide policymakers and other stakeholders with enough conceptual understanding of the variety and breadth of those alternatives to make visible that there is more than one possible path to AGI. These alternative pathways involve algorithmic and hardware technologies from diverse research areas. This suggests that policies supporting alternative technologies for AGI may need to be complex in ways that account for a variety of development entities and models.

We hope that, in this paper, we have explained some of the uncertainty around the technical path to AGI and interrogated the assumption that LLMs will simply scale to AGI. We stress this uncertainty: It is possible that LLMs could scale up in a way that compensates for their limitations and gets to transformative AGI. But there are empirical and theoretical impediments to LLM hyperscaling. Because we cannot predict the future, we urge policymakers to likewise avoid guessing the future. Although it is not in the scope of this paper to make any recommendations about specific technology or strategy, we think that U.S. government policy can account for the possibility that AGI will emerge soon in the hyperscaling paradigm without making policies that are conditioned solely on such an assumption. The U.S. government can instead plan for uncertainty and make policy choices that accommodate multiple pathways to AGI.

About the Authors

William Marcellino is a senior behavioral scientist at RAND. He develops natural language and AI tools and systems and conducts policy research on such topics as AI, foreign malign information operations, and military information technology systems. He holds a Ph.D. in rhetoric.

Lav Varshney is an associate professor of electrical and computer engineering at the University of Illinois Urbana-Champaign and an adjunct senior research engineer at RAND. His research focuses on AI theory, applications, and governance. He holds a Ph.D. in electrical engineering and computer science.

Anton Shenk is a quantitative research assistant at RAND whose focus is on the economics of emerging technologies. He has contributed to studies about AI-enhanced biological weapons and the future of economic warfare. He holds B.S. degrees in mathematics and economics.

Nicolas M. Robles is a mathematician at RAND. He designs and executes mathematical tools for defense and public policymakers and researches the impact of emerging technologies, such as AI and quantum computing in national security. He holds a Ph.D. in mathematics.

Benjamin Boudreaux is a policy researcher at RAND who works at the intersection of ethics, emerging technology, and human security. He holds a Ph.D. in philosophy.

Abbreviations

AGI	artificial general intelligence
AI	artificial intelligence
GPU	graphics processing unit
LLM	large language model
PINN	physics-informed neural network
RL	reinforcement learning

References

AlMahamid, Fadi, and Katarina Grolinger, "Autonomous Unmanned Aerial Vehicle Navigation Using Reinforcement Learning: A Systematic Review," *Engineering Applications of Artificial Intelligence*, Vol. 115, October 2022.

Aschenbrenner, Leopold, *Situational Awareness: The Decade Ahead*, June 2024.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021.

Berglund, Lukas, Meg Tong, Max Kaufmann, Mikita Balesni, Asa Cooper Stickland, Tomasz Korbak, and Owain Evans, "The Reversal Curse: LLMs Trained on 'A Is B' Fail to Learn 'B Is A,'" arXiv, arXiv:2309.12288, September 22, 2023.

Buchholz, Katharina, "The Extreme Cost of Training AI Models," *Forbes*, August 23, 2024.

Chollet, François, "On the Measure of Intelligence," arXiv, arXiv:1911.01547, November 25, 2019.

Chollet, François, "François Chollet on Deep Learning and the Meaning of Intelligence," *Mindscape* podcast, June 24, 2024.

Cuomo, Salvatore, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli, "Scientific Machine Learning Through Physics-Informed Neural Networks: Where We Are and What's Next," *Journal of Scientific Computing*, Vol. 92, No. 3, 2022.

Drexler, K. Eric, *Reframing Superintelligence: Comprehensive AI Services as General Intelligence*, Future of Humanity Institute at Oxford University, Technical Report 2019-1, 2019.

Duan, Jiafei, Samson Yu, Hui Li Tan, Hongyuan Zhu, and Cheston Tan, "A Survey of Embodied AI: From Simulators to Research Tasks," *IEEE Transactions on Emerging Topics in Computational Intelligence*, Vol. 6, No. 2, April 2022.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models," arXiv, arXiv:2303.10130, August 21, 2023.

Everitt, Tom, Ramana Kumar, Victoria Krakovna, and Shane Legg, "Modeling AGI Safety Frameworks with Causal Influence Diagrams," arXiv, arXiv:1906.08663, June 20, 2019.

Garcez, Artur S. d'Avila, Luís C. Lamb, and Dov M. Gabbay, *Neural-Symbolic Cognitive Reasoning*, Springer, 2019.

Gill, Indermit, "Whoever Leads in Artificial Intelligence in 2030 Will Rule the World Until 2100," Brookings Institution, January 17, 2020.

Goldman Sachs, *Global Macro Research: Gen AI: Too Much Spend, Too Little Benefit?* No. 129, June 25, 2024.

Golovneva, Olga, Zeyuan Allen-Zhu, Jason Weston, and Sainbayar Sukhbaatar, "Reverse Training to Nurse the Reversal Curse," arXiv, arXiv:2403.13799, May 7, 2024.

Holtman, Koen, "Counterfactual Planning in AGI Systems," arXiv, arXiv:2102.00834, January 21, 2021.

Hughes, Josie, Arsen Abdulali, Ryman Hashem, and Fumiya Iida, "Embodied Artificial Intelligence: Enabling the Next Intelligence Revolution," *IOP Conference Series: Materials Science and Engineering*, Vol. 1261, No. 1, 2022.

Ifargan, Tal, Lukas Hafner, Maor Kern, Ori Alcalay, and Roy Kishony, "Autonomous LLM-Driven Research from Data to Human-Verifiable Research Papers," arXiv, arXiv:2404.17605, April 24, 2024.

Kacianka, Severin, Amjad Ibrahim, Alexander Pretschner, Alexander Trende, and Andreas Lüdtke, "Extending Causal Models from Machines into Humans," arXiv, arXiv:1911.04869, October 31, 2019.

Korinek, Anton, "LLMs Level Up—Better, Faster, Cheaper: June 2024 Update to Section 3 of 'Generative AI for Economic Research: Use Cases and Implications for Economists,' Published in the *Journal of Economic Literature* 61(4)," American Economic Association, 2024.

Kruppa, Miles, and Deepa Seetharaman, "A Godfather of AI Just Won a Nobel. He Has Been Warning the Machines Could Take Over the World," *Wall Street Journal*, October 9, 2024.

Kurshan, Eren, "From Pursuit of the Universal AGI Architecture to Systematic Approach to Heterogeneous AGI (SAGI): Addressing Alignment, Energy & AGI Grand Challenges," *International Journal of Semantic Computing*, Vol. 18, No. 3, 2024.

Liu, Aixin, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al., "DeepSeek-V3 Technical Report," arXiv, arXiv:2412.19437, December 27, 2024.

Mercer, Sarah, Samuel Spillard, and Daniel P. Martin, "Brief Analysis of DeepSeek R1 and Its Implications for Generative AI," arXiv, arXiv:2502.02523, February 7, 2025.

Mirzadeh, Iman, Keivan Alizadeh, Hooman Shahrokhi, Oncel Tuzel, Samy Bengio, and Mehrdad Farajtabar, "GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models," arXiv, arXiv:2410.05229, October 7, 2024.

Mitchell, Melanie, "The LLM Reasoning Debate Heats Up," *AI: A Guide for Thinking Humans* blog, October 20, 2024.

Mitchell, Melanie, and David C. Krakauer, "The Debate over Understanding in AI's Large Language Models," arXiv, arXiv:2210.13966, February 10, 2023.

Morris, Meredith Ringel, Jascha Sohl-Dickstein, Noah Fiedel, Tris Warkentin, Allan Dafoe, Aleksandra Faust, Clement Farabet, and Shane Legg, "Position: Levels of AGI for Operationalizing Progress on the Path to AGI," in Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, eds., *ICML '24: Proceedings of the Forty-First International Conference on Machine Learning*, July 21, 2024.

Nayak, Anuj K., and Lav R. Varshney, "An Information Theory of Compute-Optimal Size Scaling, Emergence, and Plateaus in Language Models," submission to the International Conference on Learning Representations 2025, December 15, 2024.

Nesov, Vladimir, "Musings on LLM Scale (Jul 2024)," *Less Wrong* blog, July 3, 2024.

OpenAI, "OpenAI Charter," webpage, undated. As of November 5, 2024: https://openai.com/charter/

Pternea, Moschoula, Prerna Singh, Abir Chakraborty, Yagna Oruganti, Mirco Milletari, Sayli Bapat, and Kebei Jiang, "The RL/LLM Taxonomy Tree: Reviewing Synergies Between Reinforcement Learning and Large Language Models," arXiv, arXiv:2402.01874, February 2, 2024.

Radmehr, Bahar, Adish Singla, and Tanja Käser, "Towards Generalizable Agents in Text-Based Educational Environments: A Study of Integrating RL with LLMs," arXiv, arXiv:2404.18978, April 29, 2024.

Raissi, M., P. Perdikaris, and G. E. Karniadakis, "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations," *Journal of Computational Physics*, Vol. 378, February 2019.

Sandini, Giulio, Alessandra Sciutti, and Pietro Morasso, "Artificial Cognition vs. Artificial Intelligence for Next-Generation Autonomous Robotic Agents," *Frontiers in Computational Neuroscience*, Vol. 18, 2024.

Schaeffer, Rylan, Brando Miranda, and Sanmi Koyejo, "Are Emergent Abilities of Large Language Models a Mirage?" in A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, eds., *Proceedings of NeurIPS '23: Advances in Neural Information Processing Systems*, Vol. 36, 2024.

Schmidhuber, Juergen, "Annotated History of Modern AI and Deep Learning," arXiv, arXiv:2212.11279, December 29, 2022.

Sreedevi, A. G., T. Nitya Harshitha, Vijayan Sugumaran, and P. Shankar, "Application of Cognitive Computing in Healthcare, Cybersecurity, Big Data and IoT: A Literature Review," *Information Processing and Management*, Vol. 59, No. 2, March 2022.

Stojkovic, Jovan, Esha Choukse, Chaojie Zhang, Inigo Goiri, and Josep Torrellas, "Towards Greener LLMs: Bringing Energy-Efficiency to the Forefront of LLM Inference," arXiv, arXiv:2403.20306, March 29, 2024.

Tang, Zhengzheng, and Eva Zhu, "BrainTransformers: SNN-LLM," arXiv, arXiv:2410.14687, October 23, 2024.

Vafa, Keyon, Justin Y. Chen, Ashesh Rambachan, Jon Kleinberg, and Sendhil Mullainathan, "Evaluating the World Model Implicit in a Generative Model," arXiv, arXiv:2406.03689, November 10, 2024.

Villalobos, Pablo, Anson Ho, Jaime Sevilla, Tamay Besiroglu, Lennart Heim, and Marius Hobbhahn, "Position: Will We Run Out of Data? Limits of LLM Scaling Based on Human-Generated Data," in Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, eds., *ICML '24: Proceedings of the Forty-First International Conference on Machine Learning*, July 21, 2024.

Vöpel, Henning, "The AI Revolution: A New Paradigm of Economic Order," *Economists' Voice*, Vol. 21, No. 2, 2024.

Voss, Peter, and Mladjan Jovanovic, "Concepts Is All You Need: A More Direct Path to AGI," arXiv, arXiv:2309.01622, September 4, 2023.

Woodside, Thomas, "Emergent Abilities in Large Language Models: An Explainer," Center for Security and Emerging Technology, Georgetown University, April 16, 2024.

Yu, Haizi, James A. Evans, and Lav R. Varshney, "Information Lattice Learning," *Journal of Artificial Intelligence Research*, Vol. 77, July 2023.

Yu, Zheng, Yikuan Li, Joseph Kim, Kaixuan Huang, Yuan Luo, and Mengdi Wang, "Deep Reinforcement Learning for Cost-Effective Medical Diagnosis," arXiv, arXiv:2302.10261, February 28, 2023.

Zhang, Wanyu, and Yan Luo, "Spotlight Series on Global AI Policy—Part III: China's Policy Approach to Artificial Intelligence," *Global Policy Watch*, February 8, 2024.

Zhang, Xin, and Victor S. Sheng, "Neuro-Symbolic AI: Explainability, Challenges, and Future Trends," arXiv, arXiv:2411.04383v1, November 7, 2024.

Zhang, Xiaotian, Zhi Dou, Seung Hyun Kim, Gaurav Upadhyay, Daniel Havert, Sehong Kang, Kimia Kazemi, Kai-Yu Huang, Onur Aydin, Raymond Huang, et al., "Mind In Vitro Platforms: Versatile, Scalable, Robust, and Open Solutions to Interfacing with Living Neurons," *Advanced Science*, Vol. 11, No. 11, March 20, 2024.

Zhou, Lexin, Wout Schellaert, Fernando Martínez-Plumed, Yael Moros-Daval, Cèsar Ferri, and José Hernández-Orallo, "Larger and More Instructable Language Models Become Less Reliable," *Nature*, Vol. 634, No. 8032, October 3, 2024.