ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

The Accelerated Trajectory to AGI: A Research Assessment of Current State-of-the-Art, Bottlenecks, and Strategic Compute Dynamics

A Status Report for Machine Learning Practitioners and Data Scientists

---

Executive Summary

The trajectory toward Artificial General Intelligence (AGI)—defined as AI matching or surpassing human capabilities across virtually all cognitive tasks —is accelerating faster than predicted by general trend extrapolation, driven primarily by architectural innovations and a fundamental shift in compute utilization. Current quantitative assessments place the state of the art well past the halfway mark, generating significant confidence among industry leaders that AGI is a near-term inevitability, provided critical technical barriers are resolved.

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

Current AGI Status and Proximity

Current assessment frameworks utilizing comprehensive metrics derived from the Cattell-Horn-Carroll (CHC) theory of human intelligence confirm that frontier models have achieved an unprecedented level of capability. As of the most recent evaluation, models such as the equivalent of GPT-5 have reached approximately 57% on a comprehensive AGI score, a substantial leap from GPT-4's 27%. This places the field "roughly halfway there" to achieving human-level competence. This progress, however, presents a striking paradox: modern AI systems can perform highly complex abstract tasks, such as solving mathematical theorems and composing sophisticated poetry. Yet, these same systems fundamentally struggle with basic, human-like abilities, such as fully learning from a single example or demonstrating intuitive physics understanding necessary to "tie its shoes".

The Dual Accelerants of the AGI Timeline

The dramatic acceleration in AGI forecasting is not solely attributed to increasing model size (pre-training scaling). Instead, progress is now scaling along a "second axis" centered on post-training optimization and deployment efficiency. The two primary drivers are:

> **Post-Training Enhancements (PTE):** Techniques such as fine-tuning for tool use and strategic prompting yield significant performance increases for a fraction of the cost, quantified as a massive Compute-Equivalent Gain (CEG).
>
> **Test-Time Compute Scaling (TTC):** Methods that allow models to "think longer" during inference, enabling sophisticated, multi-step reasoning.

This operational shift has led to a compression of the doubling time for task complexity handling, accelerating from a historical rate of 210 days to approximately 135 days in 2025. Extrapolating this current pace suggests a significant probability (50% chance) of reaching 95%+ on the established AGI framework by the end of 2028, with an 80% chance by the end of 2030.

The Unsolved Barriers Defining the Race

Despite this rapid acceleration, AGI development remains critically dependent on breakthroughs in two fundamental domains where current progress is insufficient:

> **Continual Learning (CL):** This is identified as the largest unsolved problem. Existing systems exhibit "catastrophic forgetting" when learning new information, and analysis suggests long-term memory storage shows "effectively at zero progress" with no clear trajectory for forecasting.
>
> **World Modeling and Embodied Cognition:** Models lack a robust understanding of intuitive physics. On benchmarks like IntPhys 2, the best existing models perform only slightly better than chance (50%), confirming a profound gap between statistical knowledge and causal comprehension of the physical world.

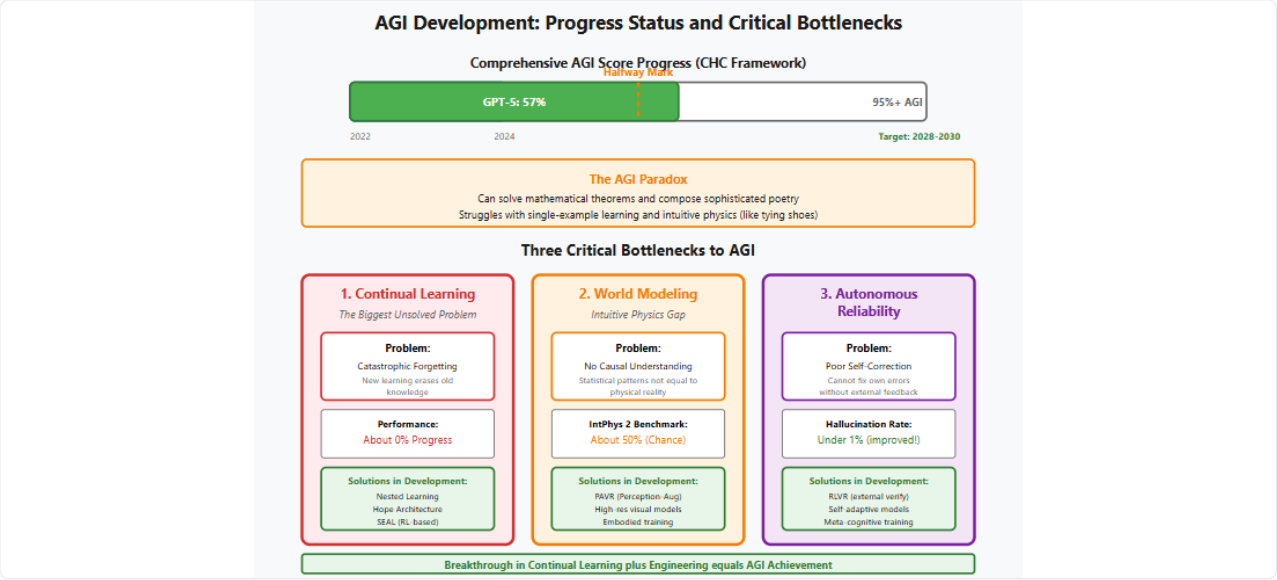
The current race to AGI is, therefore, a dual strategic challenge: maximizing the effectiveness of existing models through advanced compute deployment (TTC) while simultaneously seeking the necessary trend-bearing architectural innovations to resolve the fundamental limits of learning and memory.

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

The State of AGI: Capabilities, Metrics, and Forecast Compression

Benchmarking Human Equivalence and Current Metrics

The field of AI forecasting has moved beyond simple Turing tests, which are criticized for not accurately measuring general intelligence. Researchers now employ comprehensive frameworks, often drawing from established psychological theories of human intellect, such as the Cattell-Horn-Carroll (CHC) theory. This allows for a structured, quantitative assessment of AGI progress across dimensions including knowledge, reasoning, memory, and writing.



The evaluation metrics confirm a highly non-linear rate of progress in frontier models. For instance, the transition from GPT-4's approximate 27% score to GPT-5's assessment at 57% on a comprehensive AGI score demonstrates that capability gains are compounded, validating the perception that the technology is rapidly approaching the core cognitive thresholds required for generalized competence. This quantitative progress underpins the escalating urgency and heightened forecasts now prevalent among industry leaders.

Analyzing Timeline Compression and Expert Divergence

The most notable feature of the current AGI landscape is the sharp compression of expected timelines, coupled with a growing divergence between proprietary R&D expectations and generalized academic consensus. The median expert forecast for AGI has moved forward aggressively, shifting from expectations in the 2040s (prior to 2020) to around 2030 today. This compression is corroborated by aggregated forecasting platforms like Metaculus, which, as of late 2024, showed a median forecast of AGI arrival by 2031, representing a dramatic decrease from the multi-decade forecasts of previous years. This shift reflects the sentiment that AGI is now a "realistic near-term possibility" for approximately 30% of AI researchers.

However, the timelines proposed by industry leaders are significantly more aggressive, clustering tightly in the mid-to-late 2020s. Sam Altman of OpenAI declared confidence in knowing how to build AGI in January 2025. Similarly, Dario Amodei (Anthropic CEO) and Elon Musk project the singularity or superintelligence by 2026, while Eric Schmidt (former Google CEO) anticipates AGI within 3-5 years (i.e., 2028-2030).

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

This discrepancy between external expert trends and internal company confidence is significant. Traditional trend extrapolation suggests that achieving AGI R&D automation before 2028 would require a "reasonably large trend-breaking breakthrough" rather than simply steady progress. The high confidence expressed by CEOs is rooted in the operational success of proprietary architectural breakthroughs, specifically the ability to teach models sophisticated reasoning using reinforcement learning. This success serves as the justification for extrapolating a far shorter timeline, suggesting that the necessary breakthroughs are either imminent or already partially realized within closed research labs.

The Exponential Scaling of Utility

Further evidence of acceleration is found in the analysis of achievable task complexity, which has become a more relevant measure of utility than parameter count. The progress metric has shifted from model size to the increasing complexity and duration of tasks that models can autonomously complete.
Between 2020 and 2024, the length of software engineering and computer use tasks achievable by AI rose exponentially, growing from tasks requiring only a few seconds for a human expert to nearly an hour. If this exponential trend continues—with the most recent rate showing a doubling every four months post-2024—the models could achieve the capability to autonomously complete multi-week projects by 2028. This capacity for multi-week, unsupervised project execution represents a crucial inflection point, moving AI beyond mere chatbots into the realm of fully autonomous agents that would satisfy many definitions of AGI.
The convergence of aggressive CEO forecasts, the quantitative halfway mark on AGI score, and the exponential scaling of task complexity establishes a clear strategic mandate: the next three to five years constitute the critical window for potential AGI deployment.

Table 1: AGI Timeline Projections and Confidence Assessment (2025 Update)

Source/Expert Group	AGI Prediction	Confidence/Definition Notes	Salient Shift Rationale
CEO Consensus (Altman, Amodei, Musk, Schmidt)	2026–2028	High-confidence predictions based on internal architectural and training breakthroughs.	Driven by recent success in teaching models to reason using Reinforcement Learning.
Expert Survey Median Forecast (Metaculus, 2024)	~2031 (50% chance)	Represents aggregated community forecasting; median dramatically reduced from 50 years prior to 2020.	LLM success has made AGI a "realistic near-term possibility" for a significant minority (30%) of researchers.
Current Assessment Framework (Rachona/GitHub)	95%+ AGI Score by end of 2028 (50% chance)	Requires a large, trend-breaking breakthrough in core bottlenecks (Continual Learning).	Progress accelerating (task complexity doubling time reduced from 210 to 135 days).

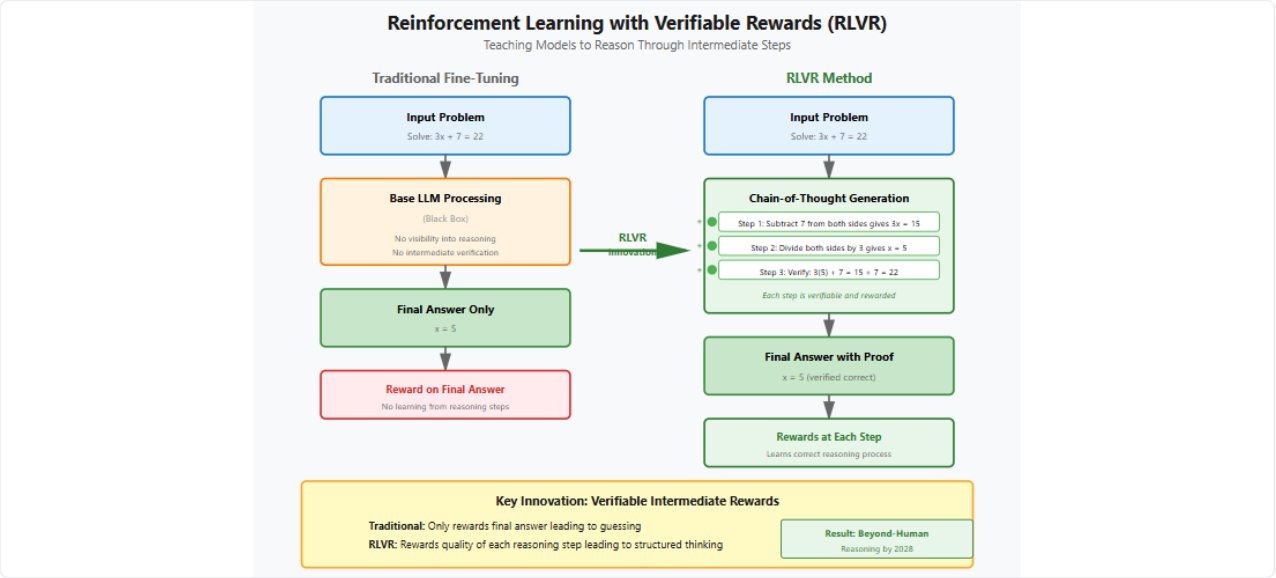
The Technical Drivers of Accelerated Reasoning and Autonomy

The field’s current velocity stems from successful strategies in maximizing model utility after the initial large-scale pre-training phase. These advancements constitute a significant scaling axis that bypasses the declining marginal returns associated with perpetually increasing model size.

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

Reinforcement Learning with Verifiable Rewards (RLVR)

A key architectural advancement driving superior reasoning capabilities is the application of Reinforcement Learning with Verifiable Rewards (RLVR). This method, utilized in models such as DeepSeek-R1 , moves beyond simply fine-tuning a model based on the final answer. Instead, RLVR optimizes the model by providing verifiable rewards linked to the quality and correctness of intermediate reasoning steps.
This framework allows the model to learn efficiently from free exploration, implicitly incentivizing the base Large Language Model (LLM) to perform correct, structured reasoning, rather than just guessing the final token. The effectiveness of RLVR is demonstrated through specialized metrics like CoT-Pass@K, which specifically measures reasoning success by evaluating both the final solution and the coherence of the intermediate Chain-of-Thought (CoT) steps. By extending the reasoning boundary for complex tasks, especially in mathematics and coding, RLVR has provided the technical impetus for the belief that beyond-human reasoning capabilities are achievable by 2028.



The Inference Compute Revolution (Test-Time Scaling)

The dramatic gains in complex reasoning are inextricably linked to a revolution in how compute is applied during inference. Test-Time Compute (TTC), sometimes referred to as "long thinking," involves allocating substantial computational resources during the prediction phase to enhance accuracy and robustness.
TTC utilizes several sophisticated methodologies, including:

- **Chain-of-Thought (CoT) Prompting:** Breaking down complex tasks into sequential, simpler steps.
- **Search and Exploration:** Evaluating multiple paths within a tree-like structure of potential responses.
- **Sampling with Majority Voting:** Generating multiple possible outputs and selecting the statistically most frequent or highest-quality answer.

While TTC dramatically improves performance on complex, multi-step problems, it is extremely resource-intensive. Analysis indicates that long thinking uses more than 100 times the compute required for a simple inference task, such as summarizing an email. The operational cost of these advanced capabilities is staggeringly high; for instance, achieving top performance on advanced benchmarks like ARC-AGI required the use of approximately 10,000 Nvidia H100 GPUs for just a 10-minute response time on a single task,

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

involving the generation of millions of tokens through parallel explorations. This reality explains why the operation of advanced systems, such as ChatGPT Pro, can run at a substantial loss, highlighting the massive financial and infrastructure backing required to operate at the cutting edge.

Agentic AI Architectures and Autonomous Tool Use

The transition from powerful foundation models to operational AGI is predicated on the rise of Agentic AI Systems. These systems represent a paradigm shift, transforming AI from passive assistants into autonomous, goal-driven decision-makers capable of end-to-end task execution.
Agentic AI relies on enhanced LLM reasoning capabilities (e.g., GPT-5, Llama 4) and is structured around three integrated layers:

Perception Layer: Responsible for collecting multimodal information from the environment, including web browsing, sensors, and APIs.

Cognition Layer: The core reasoning engine that utilizes memory and advanced planning models to analyze data, decompose goals into sequential steps, and make decisions.

Action Layer: Executes the planned steps by integrating with and manipulating external tools, such as databases, GitHub, APIs, and CRM systems.

The true utility of Agentic AI is realized through multi-agent workflows, where tasks are broken down and delegated among specialized AI tools, simulating an automated digital team that can research, write, test, and deploy content autonomously. This architectural approach is crucial for achieving the multi-week autonomous project completion forecasted by 2028.

Strategic Centralization of Operational Power

The economic dynamics of the new scaling axis reveal a critical strategic tension: the ability to enhance AI is democratized, but the ability to operate it at peak capability is centralized.
Post-training enhancements (PTEs) are relatively inexpensive to develop, often costing less than 1% of the original training expense, yet they provide substantial Compute-Equivalent Gains (CEG). This means that frontier open-source models (such as Mistral-8x22B or enhanced Llama variants) can rapidly close the performance gap with closed-source models through accessible fine-tuning techniques. This democratization facilitates the rapid proliferation of highly capable AI systems across a wide range of actors.
However, the pursuit of AGI-level performance necessitates extreme computational intensity during deployment. Test-Time Compute (TTC) is indispensable for handling the high complexity required by autonomous agents and is prohibitively expensive. The massive capital expenditure and infrastructure required to maintain this operational edge centralize the highest-performing capabilities within a small number of organizations. This structural requirement dictates that, while research and base model enhancement may diffuse widely, the deployment and scalable commercialization of the most advanced, highest-performing AGI-related agents will remain highly centralized. This dynamic is already manifesting in the enterprise market, where corporate spending is consolidating around a few high-performing, closed-source models, resulting in market share shifts, such as Anthropic gaining ground on OpenAI. The ability to "think longer" through massive inference compute has become the new competitive differentiator, establishing inference capacity as the controlling strategic resource.

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

The Unsolved Barriers: Continual Learning and Memory Systems

Despite the acceleration in reasoning and autonomy, the most critical AGI bottleneck remains the inability of current architectures to learn organically over time without corrupting prior knowledge.

Catastrophic Forgetting (CF): The Core Developmental Blockade

Continual Learning (CL) is widely recognized as the single biggest unsolved problem. Existing LLMs function as static, fixed-knowledge systems whose abilities are confined to the information learned during pre-training or the immediate context window. When these models are fine-tuned on new domain-specific data, they suffer from catastrophic forgetting, a phenomenon where they lose general knowledge, logical reasoning abilities, and competence in tasks learned previously.

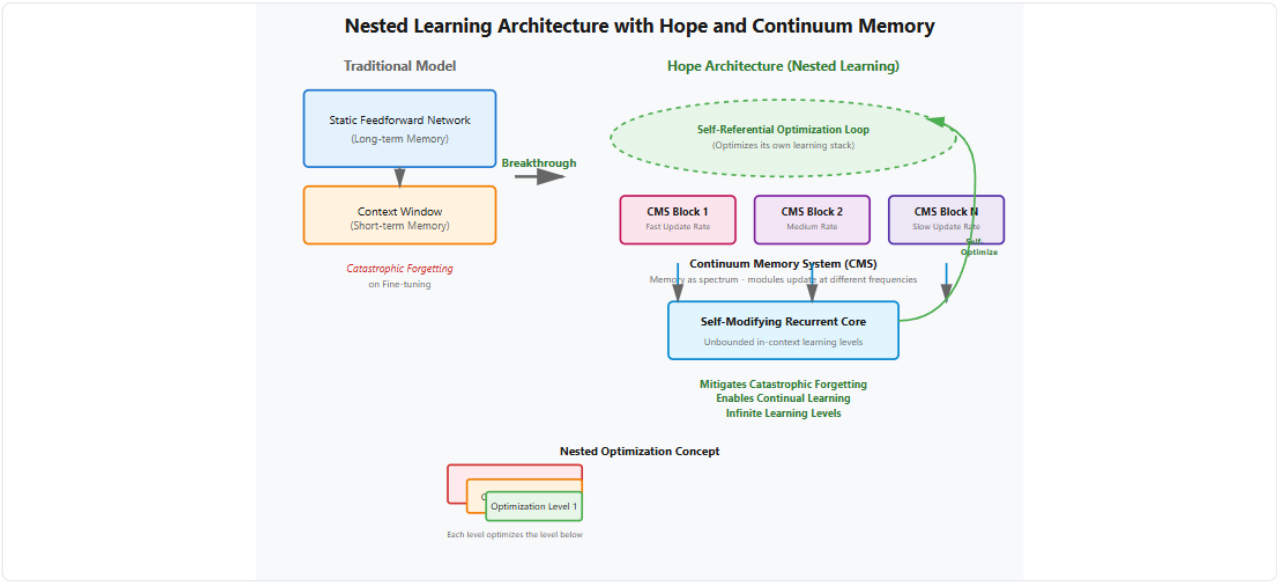
This deficit prevents LLMs from exhibiting fundamental cognitive functions of AGI: they cannot build context, analyze their failures, or adapt to new information organically through feedback and experience. Crucially, in terms of long-term memory storage, the field is assessed as being "effectively at zero progress" with no measurable trajectory forward, making forecasting in this domain impossible based on extrapolation from performance curves. A breakthrough in CL, alongside standard engineering improvements in other areas, may be the sole requirement remaining for AGI attainment.

Novel Memory Architectures to Mitigate CF

Leading research is focused on architectural modifications that enable neuroplasticity, drawing inspiration from biological systems where forgetting is gradual and memory consolidation is active.
Google Research's introduction of Nested Learning (NL) represents a significant conceptual leap. NL views the machine learning process not as a single fixed optimization problem, but as a set of smaller, nested optimization problems, thereby bridging the gap between static network architecture and the optimization algorithm. This paradigm aims to mitigate or completely avoid catastrophic forgetting.

The NL framework extends memory into a Continuum Memory System (CMS). In standard transformer models, short-term memory is the context window, and long-term memory is the static feedforward network. CMS, however, treats memory as a spectrum of modules, each updating at different, specific frequency rates. The practical realization of NL principles is the Hope architecture. Hope is a self-modifying recurrent architecture augmented with CMS blocks. It extends previous concepts (like Titans architectures, which prioritize memory based on surprise) by allowing for "unbounded levels of in-context learning" through a self-referential process. This design means the system can optimize its own learning stack and memory parameters on the fly, creating an architecture capable of infinite, looped learning levels. This architectural shift, enabling the stack itself to optimize and scale with compute, is recognized as the essential requirement to surpass limitations imposed by manual engineering and resolve the static memory problem.

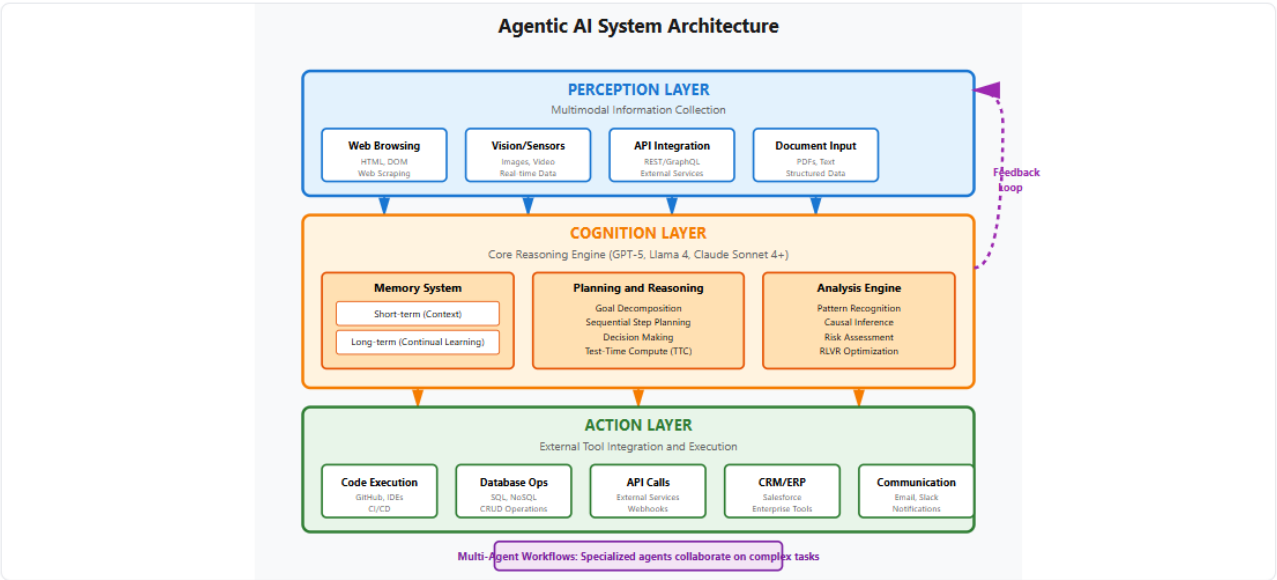
ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk



Agentic Memory Management and Experience Replay

For autonomous agents, sophisticated memory management is required to leverage historical experiences effectively. Current memory systems for LLM agents are moving beyond basic graph databases to dynamic, structured systems. Novel proposals, drawing organizational principles from methods like the Zettelkasten knowledge-management technique, aim to create interconnected knowledge networks through dynamic indexing and linking.

A critical feature of these advanced agentic systems is memory evolution. As new data is integrated, it can trigger updates to the contextual representations and attributes of existing historical memories, allowing the entire memory network to continuously refine its understanding. However, this complexity introduces new challenges, notably error propagation and misaligned experience replay, where past faulty decisions or corrupted memories negatively impact future agent behavior. The reliability of the evaluator signal is therefore paramount for effective agentic memory management.



ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

Autonomous Self-Adaptation Frameworks

Beyond memory retention, research focuses on enabling autonomous self-improvement. Self-adaptive LLMs must dynamically learn and adjust their weights in response to new tasks, eliminating the need for computationally intensive, static fine-tuning cycles. Frameworks are leveraging Reinforcement Learning (RL) to facilitate this lasting adaptation. For example, SEAL (Self-Edits via Reinforcement Learning) uses an RL loop to train the model to produce persistent weight updates, or "self-edits," using the updated model's downstream performance as the reward signal.

Unlike prior approaches relying on separate adaptation modules, SEAL parameterizes and controls its own adaptation process using its existing generative capabilities, representing a significant step toward achieving true long-term autonomous self-improvement. The collective movement toward Nested Learning, Continuum Memory Systems, and RL-driven self-editing highlights a fundamental understanding: AGI requires a system defined by dynamism and neuroplasticity. The ability to modify one’s own architecture and learning process is the prerequisite for achieving continual, unbounded learning.

The Perception and World Modeling Challenge

While the field rapidly masters abstract reasoning, a parallel and equally critical challenge involves grounding AI systems in the physical, causal reality of the world. This is encapsulated by the persistent deficits in intuitive physics and fine-grained perception.

Deficits in Intuitive Physics and World Modeling

Intuitive physics understanding—the ability to predict how objects will interact and behave in the real world—remains a major gap between current AI and AGI. The IntPhys 2 benchmark, designed to test this fundamental capability across complex synthetic environments, starkly illustrates this shortfall. State-of-the-art models perform only marginally better than chance (around 50%) on these tasks, failing to grasp basic causality, which stands in sharp contrast to the near-perfect accuracy achieved by human performance. This deficit confirms that current AI, despite its intellectual prowess in symbolic manipulation, lacks a foundational causal world model.

Multimodal Integration and the Perception Bottleneck

The advancement of Multimodal Large Language Models (MLLMs), such as GPT-4 Vision, has demonstrated strong end-to-end embodied decision-making abilities. However, detailed analysis of Abstract Visual Reasoning (AVR) tasks indicates that the core limitation lies not in the high-level reasoning capacity of the LLM component, but in a deficiency of fine-grained perception.

The failure to maintain robust performance when visual logic puzzles are simply enlarged underscores this issue: while humans barely notice the difference, model performance degrades significantly. This suggests a fundamental perception bottleneck, rather than a failure of reasoning itself, driving the observed limitations. To address this, cutting-edge frameworks like the Perception-Augmented Visual Reasoner (PAVR) have been developed. PAVR emphasizes a two-stage training paradigm that progressively enhances the perceptual ability before strengthening the reasoning component. This approach has proven substantially superior to both open-source and commercial MLLMs, confirming fine-grained perception as the primary challenge in abstract visual tasks. Furthermore, new MLLMs (e.g., InternLM-XComposer2-4KHD, Qwen2-VL) are

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

incorporating high-resolution visual perception mechanisms to build a more robust foundation for visual understanding.

The Statistical vs. Embodied Intelligence Gap

Even with multimodal capabilities, advanced LLMs continue to exhibit a deficit in true sensory grounding, which is essential for human-like cognition. While top models can achieve 85%–90% accuracy in approximating human sensory-linguistic associations and strong correlation with human ratings, they still differ from human embodied cognition in their processing mechanisms. The analysis suggests that introducing multimodality often provides similar statistical information as training on massive text alone, rather than qualitatively distinct, sensorily grounded data.

The simultaneous achievement of high scores on abstract CHC intelligence metrics (57% AGI score) and near-chance performance on intuitive physics benchmarks (IntPhys 2) confirms a critical finding: current AI systems excel at statistical pattern recognition and abstract inference but fundamentally lack genuine sensory grounding and coherent causal world models. A shift from purely statistical representation to true embodied, causally coherent processing of sensory data is necessary to bridge this intelligence gap.

Reliability, Hallucinations, and Self-Correction

Significant engineering effort has substantially improved model reliability. Top models now exhibit hallucination rates of less than 1%, a major reduction from the 15%–20% rate observed just two years prior. However, even advanced LLMs struggle with **intrinsic self-correction**. Research indicates that when models are asked to detect and fix their own errors internally, without external feedback (such as verified final outcomes), their performance struggles, and in some cases, can even degrade. This limitation underscores the reliance of current systems on external verification mechanisms (like RLVR) to ensure accuracy in complex, multi-step planning, rather than possessing the internal metacognitive ability to analyze and correct their own fundamental failures.

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

Table 2: Critical AGI Bottlenecks and Cutting-Edge Mitigation Strategies

AGI Bottleneck	Core Limitation	Measured Performance Benchmark	Cutting-Edge Solution/Framework
Continual Learning/Memory	Catastrophic Forgetting; Static knowledge base; "Zero progress" in long-term memory trajectory.	Inability to build context or adapt to new information organically.	Nested Learning / Hope Architecture ; SEAL (Self-Edits via RL) ; Agentic Memory Systems (dynamic organization).
World Modeling/Intuitive Physics	Lack of causal understanding; Failure to predict physical reality.	IntPhys 2 models perform at chance levels (50%).	Perception-Augmented Visual Reasoner (PAVR) ; High-resolution visual perception mechanisms (e.g., Qwen2-VL).
Autonomous Execution Reliability	Error propagation in multi-step planning; Lack of robust, intrinsic self-correction.	LLMs struggle to self-correct without external feedback, sometimes degrading performance.	Reinforcement Learning with Verifiable Rewards (RLVR) ; Self-adaptive reasoning models (e.g., CLIO).

Strategic Implications, Economic Shifts, and Governance

The fundamental technical shifts toward post-training scaling and extreme inference compute have profound implications for the competitive landscape, economic structures, and geopolitical governance of AGI development.

The Compute-Equivalent Gain (CEG) as the New Competitive Metric

The rise of post-training enhancements (PTEs) provides enormous strategic leverage. PTEs—including tool-use fine-tuning, advanced prompting, and scaffolding—yield substantial performance improvements, which can be translated into a Compute-Equivalent Gain (CEG). Quantifying this effect shows that most surveyed enhancements improve benchmark performance by more than a 5\times increase in training compute, with some enhancements exceeding a 20\times gain.
This metric validates the industrial focus on fine-tuning and post-training optimization, offering a strategic reprieve from the diminishing returns of perpetual pre-training scaling. The implication is that competitive advantage is increasingly derived not solely from the initial, massive training run, but from the highly specialized and low-cost development of PTEs.

Market Dynamics: Open Source Diffusion vs. Enterprise Consolidation

The difference in costs and capabilities creates a dual market reality. The low cost and high CEG associated with enhancement techniques facilitate the rapid diffusion of powerful models. Open-source models, which are often only a few months behind the frontier, can be copied, fine-tuned, and run on high-end consumer hardware. This democratization ensures that AI capabilities spread widely and cheaply.
Conversely, enterprise market dynamics show a consolidation of spending around models that offer peak reliability and performance, which often rely on proprietary, closed-source architectures and sophisticated Test-Time Compute (TTC) infrastructure. For example, Anthropic has surpassed competitors in enterprise usage due to performance advantages, signifying a consolidation of enterprise dollars around closed models capable of achieving high operational standards. The operational difficulty and extreme cost of running the most complex autonomous agents (TTC) mean that while capability diffuses, true, reliable, frontier *operational power* remains highly centralized among organizations with bespoke GPU access.

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

Economic Impact of Autonomous Labor and Scarcity

The long-term economic trajectory is clear: AGI portends a significant decline in the role of human labor, necessitating a fundamental societal and economic reevaluation. However, this transition will be mediated by several factors, including:

Production and Diffusion Lags: The time required to deploy advanced AI systems across all sectors.

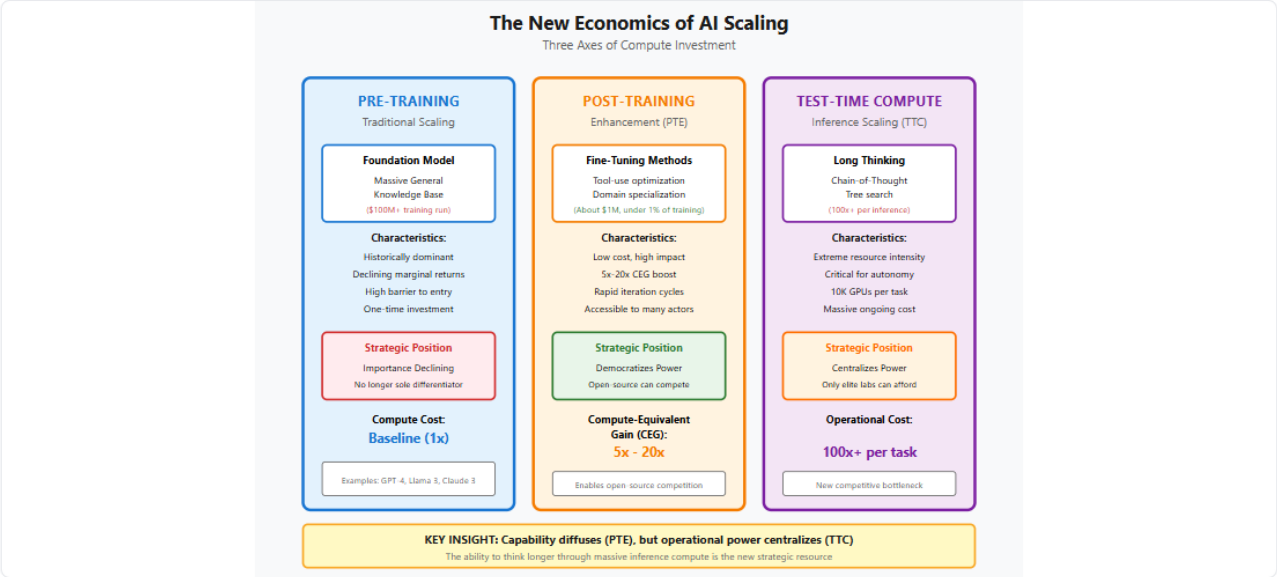
Trust Issues: Initial reluctance to rely on autonomous AI for critical decisions.

Irreproducible Human Elements: Continuing demand for authentic human connection, identity in competitive roles (sports, arts), and specialized tasks requiring implicit knowledge transfer.

The advent of AGI will dramatically increase abundance, particularly for reproducible factors like compute and robots. However, it will not create a true post-scarcity economy. Core economic trade-offs will persist. Specifically, irreproducible factors—such as land, raw materials, and energy access—are expected to appreciate in relative value compared to reproducible factors like computer chips and automated labor. The distribution and use of compute and automated production will still be subject to scarcity constraints, and these trade-offs will become central to future economic policy.

Governance and the Compute/Control Paradox

The accelerating progress and the potential for technological self-improvement necessitate immediate attention to existential risk (x-risk) from Artificial Superintelligence (ASI). Leading AI researchers and CEOs acknowledge the risk that inability to control superintelligent AI could lead to global catastrophe, declaring mitigation of extinction risk from AI a global priority. The structure of the scaling economics creates a governance paradox. On one hand, the substantial Compute-Equivalent Gain (CEG) obtained through low-cost fine-tuning makes it challenging to govern AGI development through model access controls alone, as frontier models can be enhanced by a wide range of actors. The rapid proliferation of powerful open-source variants limits the effectiveness of centralized control efforts.



ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

On the other hand, the extreme compute requirements for operationalizing the most complex, AGI-adjacent capabilities (TTC) offer a potential point of leverage. Since running these frontier models requires bespoke,

massive, and expensive GPU infrastructure (as exemplified by the 10,000 H100s needed for one complex task) , regulatory focus targeting the physical supply chain and operational capacity of high-cost inference infrastructure may prove more effective than efforts to control model weights or source code. The centralization of the operational edge provides a defined, although temporary, point of strategic control.

Table 3: The New Economics of AI Scaling (Post-Training Compute)

Scaling Strategy	Impact on Capability	Compute Consumption	Strategic Implication
Pre-Training Scaling (Traditional)	Equips models with massive general knowledge (Foundation Model).	Historically dominant; declining marginal returns now observed.	High initial barrier to entry; focus shifting away from pure size.
Post-Training Enhancements (CEG)	Provides performance gain equivalent to 5\times to 20\times additional training compute.	Costs typically <1\% of original training cost.	Democratizes performance improvement; complicates governance due to wide actor engagement.
Test-Time Compute (TTC/Long Thinking)	Enables complex, multi-step, multi-path reasoning critical for agent autonomy.	Uses 100\times the compute of simple inference; extremely resource-intensive operation (e.g., 10,000 H100s for one task).	Centralizes operational power in top labs; inference cost becomes the new competitive bottleneck.

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

Conclusions and Strategic Recommendations

The current assessment places AI firmly on an accelerated, non-linear path toward AGI, validated by quantitative metrics and profound shifts in compute strategy. The competitive race is no longer solely about the magnitude of the initial training investment but about the ability to innovate and deploy sophisticated post-training techniques effectively.

Near-Term Milestones and Strategic Window (2028 Forecast)

The next three years are critical. Extrapolation of the exponential scaling of task complexity suggests a high probability that frontier agentic models will achieve the capacity for autonomous, multi-week project completion by 2028. This milestone represents the functional tipping point for AGI, as it would likely enable the self-automation of AI research and development, initiating the rapid feedback loop often associated with the technological singularity. This operational capability would strongly correlate with the high-end expert projection of 95%+ AGI readiness by 2028.

Targeted R&D Recommendations for AGI Completion

Achieving AGI readiness hinges on specific architectural breakthroughs, not merely continued scale:

> **Priority 1: Architectural Breakthroughs in Continual Learning.** The strategic imperative is to resolve the fundamental limits of learning and memory. Investment must be prioritized in the development and scaling of dynamic, self-modifying architectures, such as Nested Learning and the Hope architecture. These systems, which allow for the optimization of the learning stack itself, represent the most promising path toward eliminating Catastrophic Forgetting, which remains the
>
> single most unpredictable variable in AGI forecasting. Furthermore, research into RL-driven self-adaptation (SEAL) is necessary to ensure lasting, persistent weight updates in autonomous agents.
>
> **Priority 2: Achieving Embodied Cognition and Causal Modeling.** A dedicated R&D focus is required to address the robust perception bottleneck in Multimodal LLMs (e.g., the PAVR approach). Success requires training models to consistently surpass chance levels on intuitive physics benchmarks (IntPhys 2), ensuring that AI development shifts definitively from statistical correlation to genuine causal understanding and sensory grounding.

Governance Recommendations

In light of the Compute/Control Paradox, strategic governance efforts must adapt to the new economic reality:

- **Regulate Operational Capacity:** Since the highest-risk, most complex autonomous capabilities require extreme Test-Time Compute (TTC) and specialized, high-cost infrastructure, policy and strategic foresight should focus on the provision, allocation, and oversight of this critical inference compute infrastructure (data centers, GPU supply chains).

- **Acknowledge Diffused Capability:** The low cost of post-training enhancements (CEG) ensures that powerful models will proliferate widely. Governance must account for the rapid diffusion of capable open-source models and focus on safety standards, verification mechanisms (like RLVR), and ethical deployment practices applicable across a broad spectrum of actors.

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

Quoted works and sources

Artificial general intelligence - Wikipedia, https://en.wikipedia.org/wiki/Artificial_general_intelligence

2025 Mid-Year LLM Market Update: Foundation Model Landscape + Economics, https://menlovc.com/perspective/2025-mid-year-llm-market-update/

AI capabilities can be significantly improved without expensive retraining - arXiv, https://arxiv.org/html/2312.07413v1

How Scaling Laws Drive Smarter, More Powerful AI - NVIDIA Blog, https://blogs.nvidia.com/blog/ai-scaling-laws/

How to Alleviate Catastrophic Forgetting in LLMs Finetuning? Hierarchical Layer-Wise and Element-Wise Regularization - arXiv, https://arxiv.org/html/2501.13669v2

IntPhys 2: Benchmarking Intuitive Physics Understanding In Complex Synthetic Environments | Research - AI at Meta, https://ai.meta.com/research/publications/intphys-2-benchmarking-intuitive-physics-understanding-in-complex-synthetic-environments/

Artificial General Intelligence: Jumping to the New Inference Market S-Curve - CMS Wire, https://www.cmswire.com/digital-experience/artificial-general-intelligence-jumping-to-the-new-inference-market-s-curve/

Shrinking AGI timelines: a review of expert forecasts - 80,000 Hours, https://80000hours.org/2025/03/when-do-experts-expect-agi-to-arrive/

When Will AGI/Singularity Happen? 8,590 Predictions Analyzed - Research AIMultiple, https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/

The case for AGI by 2030 - 80,000 Hours, https://80000hours.org/agi/guide/when-will-agi-arrive/

Reinforcement Learning with Verifiable Rewards Implicitly Incentivizes Correct Reasoning in Base LLMs - arXiv, https://arxiv.org/html/2506.14245v2

Why AI's next phase will likely demand more computational power, not less - Deloitte, https://www.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2026/compute-power-ai.html

When AI Takes Time to Think: Implications of Test-Time Compute - RAND, https://www.rand.org/pubs/commentary/2025/03/when-ai-takes-time-to-think-implications-of-test-time.html

Agentic AI Systems: The Future of Autonomous Intelligence in 2025 | by Sayali Kumbhar | Nov, 2025, https://medium.com/@sayalisureshkumbhar/agentic-ai-systems-the-future-of-autonomous-intelligence-in-2025-5ed4579fb108

Top 10 Resources to Learn Agentic AI in 2025 | by javinpaul | Javarevisited | Sep, 2025, https://medium.com/javarevisited/top-10-resources-to-learn-agentic-ai-in-2025-cef1958d9e59

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

Top 10 open source LLMs for 2025 - NetApp Instaclustr, https://www.instaclustr.com/education/open-source-ai/top-10-open-source-llms-for-2025/

Three Shaky Assumptions Underpinning many AGI Predictions : r/ControlProblem - Reddit, https://www.reddit.com/r/ControlProblem/comments/1o2n14g/three_shaky_assumptions_underpinning_many_agi/

Introducing Nested Learning: A new ML paradigm for continual learning - Google Research, https://research.google/blog/introducing-nested-learning-a-new-ml-paradigm-for-continual-learning/

Catastrophic Forgetting: The Essential Guide | Nightfall AI Security 101, https://www.nightfall.ai/ai-security-101/catastrophic-forgetting

Introducing Nested Learning: A new ML paradigm for continual learning - Reddit, https://www.reddit.com/r/singularity/comments/1or265r/google_introducing_nested_learning_a_new_ml/

A-MEM: Agentic Memory for LLM Agents - arXiv, https://arxiv.org/pdf/2502.12110

How Memory Management Impacts LLM Agents: An Empirical Study of Experience-Following Behavior - arXiv, https://arxiv.org/html/2505.16067v2

Self-Adaptive LLMs - Medium, https://medium.com/@bijit211987/self-adaptive-llms-422196f6b123

Self-Adapting Language Models - arXiv, https://arxiv.org/html/2506.10943v1

Facebookresearch/IntPhys2: This is the code repository for IntPhys 2, a video benchmark designed to evaluate the intuitive physics understanding of deep learning models. - GitHub, https://github.com/facebookresearch/IntPhys2

Towards End-to-End Embodied Decision Making via Multi-modal Large Language Model: Explorations with GPT4-Vision and Beyond - arXiv, https://arxiv.org/abs/2310.02071

VisuRiddles: Fine-grained Perception is a Primary Bottleneck for Multimodal Large Language Models in Abstract Visual Reasoning | OpenReview, https://openreview.net/forum?id=fGRwRnDVMX

VisuRiddles: Fine-grained Perception is a Primary Bottleneck for Multimodal Large Language Models in Abstract Visual Reasoning - arXiv, https://arxiv.org/html/2506.02537v3

Exploring Multimodal Perception in Large Language Models Through Perceptual Strength Ratings - arXiv, https://arxiv.org/abs/2503.06980

Large Language Models Cannot Self-Correct Reasoning Yet | OpenReview, https://openreview.net/forum?id=IkmD3fKBPQ

How DeepSeek has changed artificial intelligence and what it means for Europe - Bruegel, https://www.bruegel.org/policy-brief/how-deepseek-has-changed-artificial-intelligence-and-what-it-means-europe

The Age of Agi: The Upsides and Challenges of Superintelligence, https://www.aei.org/articles/the-age-of-agi-the-upsides-and-challenges-of-superintelligence/

ATOMIX – Where Technology and Vision Converge
Gammel Havnevej 5
4070 Kirke Hyllinge
T: +45 26176088
E: contact@atomix.dk
W: atomix.dk

Existential risk from artificial intelligence - Wikipedia, https://en.wikipedia.org/wiki/Existential_risk_from_artificial_intelligence

AGI and Democracy - Ash Center - Harvard University, https://ash.harvard.edu/wp-content/uploads/2024/03/340307-Ash-AGI-Pascal-V2-.pdf