REVIEW ARTICLE

JAIKE

The Ascent of Machine Cognition: A Systematic Review of Large Reasoning Models

Mohit Sewak
Google Inc, India

DOI:10.37819/Jaike.25.0010

Article history:
Received : 09.09.2025
Revised : 19.10.2025
Accept : 11.11.2025
Published : 05.08.2026

Corresponding author:
E-mail: mohitsewak@google.com

Abstract: The evolution of Large Language Models (LLMs) has catalyzed significant advancements in artificial intelligence; however, their initial proficiency in fluent text generation often masked an underlying fragility in tasks demanding complex, multi-step reasoning. This review charts the systematic development of Large Reasoning Models (LRMs)—a functional classification characterized not by novel architectures, but by the augmentation of existing LLMs with sophisticated techniques designed to elicit, structure, and ground their latent cognitive abilities. We synthesize this burgeoning field across four principal vectors of innovation. The first is the *elicitation of latent reasoning*, a paradigm shift initiated by Chain-of-Thought (CoT) prompting, which revealed that structured interaction could unlock inherent problem-solving capabilities. The second vector involves the development of *advanced reasoning architectures*, transitioning from linear processes to deliberate, nonlinear frameworks such as Tree of Thoughts (ToT) and Graph of Thoughts (GoT), enabling exploration, backtracking, and synthesis. The third addresses the inherent limitations of LLMs through *tool augmentation and program-aided reasoning*, grounding models in external realities by offloading precise computation and information retrieval to external systems, and extending capabilities into formal domains and autonomous agency. The fourth vector confronts the critical frontier of *self-correction and critique*, a nascent area focused on imbuing models with the capacity for introspection and iterative refinement. By deconstructing these advancements and synthesizing them in comparative tables, we present a cohesive narrative of a field moving from stochastic pattern completion toward more robust, verifiable, and adaptable machine reasoning, and conclude with a critical evaluation of the profound open challenges that will define the future trajectory of artificial intelligence.

Keywords: Large language models, machine reasoning, chain-of-thought, cognitive architectures, tool augmentation, self-correction, artificial intelligence, agentic AI

1. INTRODUCTION

The recent history of artificial intelligence has been dominated by the scaling of Large Language Models (LLMs), predominantly based on the Transformer architecture (Vaswani et al., 2017). The strategy of increasing parameter counts, computational budgets, and training data volume has yielded remarkable results. Models such as the GPT series (Radford et al., 2018; Brown et al., 2020) have demonstrated unprecedented capabilities in generating fluent, coherent, and contextually nuanced text. A significant phenomenon observed during this era is the emergence of abilities that are not explicitly programmed but arise as a function of scale (Wei et al., 2022a).

However, this linguistic prowess often belied a fundamental weakness. When tasked with problems

requiring rigorous, multi-step reasoning—such as arithmetic, symbolic logic, or complex planning—foundational LLMs frequently produced plausible-sounding but logically flawed outputs. Their operation resembled sophisticated pattern matching or "stochastic parroting" (Bender et al., 2021) rather than deliberate, verifiable problem-solving. This critical gap between fluency and logical fidelity catalyzed a profound shift in research focus: moving beyond mere scaling to the development of sophisticated techniques designed to guide and structure the "cognitive processes" of these models.

It is crucial to distinguish between the computational process of the model and the cognitive outcomes it produces. Computationally, an LLM performs *inference*—a forward pass through its neural architecture to predict the next token. In contrast, *reasoning*, in the context of this review, refers to the structured, logical, and multi-step process of solving a problem or arriving at a conclusion. The core advancement of this field lies in developing methodologies that utilize the mechanical process of inference to elicit and scaffold the cognitive process of reasoning. This review systematically charts the emergence of what we term **Large Reasoning Models (LRMs).** This designation is not architectural but *functional*. An LRM is an LLM augmented with a scaffolding of techniques—including specialized prompting strategies, complex data structures for thought exploration, interfaces for external tool use, and iterative refinement loops—that collectively enable it to transcend basic pattern completion and engage in complex reasoning. The central thesis of this review is that the evolution of LRMs represents a systematic effort to identify the inherent limitations of a core technology and iteratively construct a suite of guiding mechanisms to overcome them.

We synthesize the burgeoning literature into four distinct yet interconnected pillars of innovation that define the field's progression, summarized in Table 1 and visualized conceptually in Figure 1:

1. Eliciting Latent Reasoning (Section 2): The discovery that the reasoning capabilities of LLMs, while dormant, can be activated through structured interaction, exemplified by the seminal Chain-of-Thought approach (Wei et al., 2022b).

2. Structuring Complex Cognition (Section 3): The evolution from linear reasoning paths to sophisticated, non-linear structures like trees (Yao et al., 2023) and graphs (Besta et al., 2023), which better emulate human deliberation and exploration.

3. Grounding Reasoning in External Reality (Section 4): The crucial integration of external tools and programs (Gao et al., 2022; Schick et al., 2023) to offload tasks for which LLMs are ill-suited, thereby grounding their reasoning in verifiable, external contexts, formal systems, and enabling the rise of autonomous agents (Park et al., 2023).

4. The Frontier of Self-Correction (Section 5): The nascent yet critical pursuit of imbuing models with the ability to evaluate, critique, and iteratively refine their own outputs (Madaan et al., 2023), a key step toward autonomous and reliable systems.

By tracing this intellectual arc, this review provides a comprehensive synthesis of the state of- the-art in machine reasoning, concluding with a discussion of the significant open challenges and future research directions that will shape the development of more capable and responsible artificial intelligence.

2. THE ACTIVATION OF LATENT COGNITION: ELICITING REASONING IN LLMS

The trajectory toward LRMs began not with architectural modification, but with a paradigm shift in human-model interaction. Researchers discovered that the vast, implicit knowledge embedded within scaled LLMs (Brown et al., 2020) contained latent reasoning capabilities that remained inaccessible through standard prompting techniques. The key insight was that *how* a model is prompted is as crucial as the knowledge it contains. This marked a transition from simply requesting an answer to instructing the model to articulate its process.

2.1. The Chain-of-Thought Catalyst

The seminal development in this area was the introduction of Chain-of-Thought (CoT) prompting (Wei et al., 2022b). Prior to this, standard prompting involved presenting the model with an input (the question) and expecting a direct output (the answer). This direct input-to-output mapping proved brittle for tasks requiring sequential deduction or calculation. Wei et al. demonstrated that by providing a few examples (few-shot prompting) where the solution included the intermediate reasoning steps—a "chain of thought"—the model could adopt this strategy for novel problems.

JAIKE https://doi.org/10.37819/Jaike.25.0010

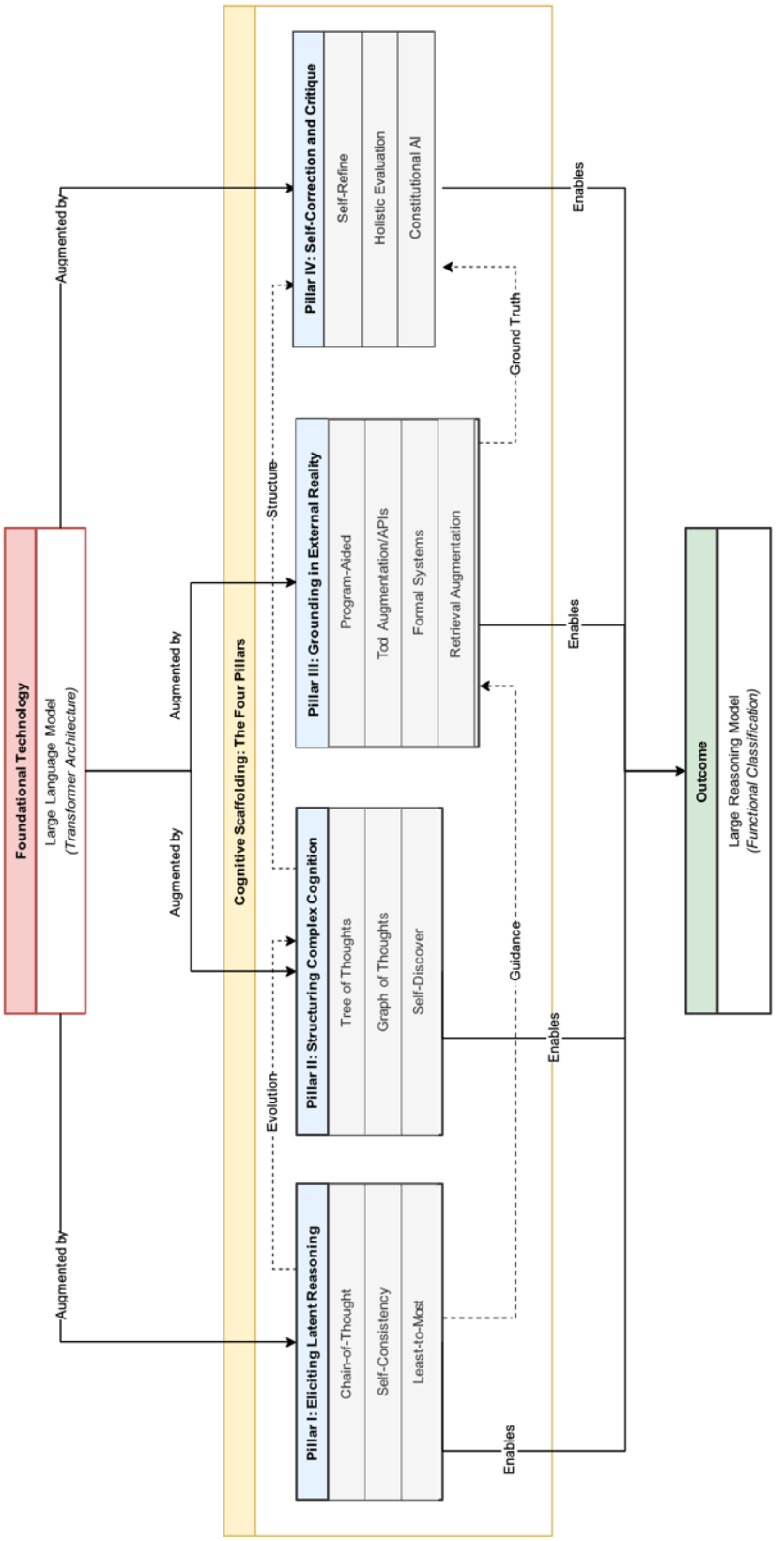

Figure 1. Conceptual framework of Large Reasoning Models (LRMs). Foundational LLMs are augmented by four interconnected pillars of innovation (cognitive scaffolding) to enable complex reasoning capabilities. Dashed lines indicate key conceptual interconnections

The impact of this intervention was significant. On benchmarks requiring mathematical and commonsense reasoning (e.g., GSM8K), CoT prompting dramatically improved performance; for instance, it boosted the accuracy of the 540B parameter PaLM model on GSM8K from 18% (standard prompting) to 57% (Wei et al., 2022b). The significance of CoT is twofold. First, it established that the reasoning process itself is a crucial output, serving as a cognitive scaffold that allows the model to decompose complexity and mitigate the risk of premature, incorrect conclusions. Second, the externalized reasoning chain offered a novel degree of interpretability, allowing researchers to diagnose *where* a model's logic faltered, rather than just observing the final error.

2.2. Generalization and Robustness

The initial formulation of CoT required the manual creation of task-specific examples. This limitation was rapidly addressed by two critical innovations that enhanced the technique's generality and robustness.

Kojima et al. (2022) introduced "Zero-Shot-CoT," discovering that reasoning could be elicited without *any* examples. By simply appending a task-agnostic trigger phrase, such as "*Let's think step by step*," they induced the model to generate a coherent chain of thought. This finding was crucial, as it suggested that the reasoning abilities elicited by few-shot CoT were not merely a result of pattern mimicry but were an emergent property of scale, inherent to the model and awaiting the appropriate cognitive trigger.

To address the fragility of individual reasoning paths, Wang et al. (2022) introduced the concept of "Self-Consistency." They observed that a single, greedily decoded chain of thought could be derailed by an isolated logical error. Their solution was an ensemble-like decoding strategy: generate multiple, diverse reasoning paths for the same problem by sampling from the model's output distribution (e.g., using temperature sampling), and then select the final answer by majority vote. This approach of marginalizing out flawed reasoning through consensus substantially improved accuracy across a spectrum of benchmarks, yielding further significant gains (e.g., improving PaLM-540B accuracy on GSM8K from 57% to 74.4%).

2.3. Decomposing Complexity

For problems characterized by significant compositional complexity, generating a correct, monolithic chain of thought can remain intractable. Addressing this, Zhou et al. (2022) introduced "Least-to-Most Prompting." This strategy involves a multi-stage process. First, the model is prompted to decompose a complex problem into a sequence of simpler, ordered subproblems.

Subsequently, the model solves these subproblems sequentially, with the crucial feature that the solution to each preceding subproblem is injected into the context for the next. This methodical decomposition

Table 1. Overview of the Four Pillars of Innovation in Large Reasoning Models (LRMs)

Pillar	Focus Area	Key Challenges Addressed	Impact on LRM Development
I	Eliciting Latent Reasoning	Inability of standard prompting to handle multistep logic; fragility of direct answers.	Established methods to unlock inherent problem-solving capabilities and improve interpretability.
II	Structuring Complex Cognition	Limitations of linear reasoning paths for tasks requiring exploration, planning, and synthesis.	Enabled deliberate, multi-path problem-solving that mirrors human deliberation and meta-cognition.
III	Grounding in External Reality	Inaccurate computation; factual hallucination; static knowledge cutoffs; lack of agency.	Grounded outputs in verifiable contexts, enabled precise calculation, facilitated formal reasoning, and enabled autonomous agency.
IV	Self-Correction and Critique	Unreliability of outputs; inability of models to autonomously identify and correct their own errors.	Pushing towards more autonomous, reliable, and trustworthy reasoning systems through iterative refinement.

JAIKE https://doi.org/10.37819/Jaike.25.0010

allows the model to manage cognitive load and tackle challenges that would be overwhelming if approached directly, proving especially effective for tasks requiring generalization far beyond the training data patterns.

Collectively, these foundational techniques transformed the perspective on LLMs, establishing that unlocking machine reasoning was not solely a matter of scale, but of designing interactions that effectively scaffold and guide their latent cognitive abilities. These methods are summarized and compared in Table 2.

3. BEYOND LINEARITY: ADVANCED STRUCTURES FOR DELIBERATE REASONING

The success of the linear chain of thought prompted researchers to explore more sophisticated reasoning structures that better emulate the nuances of human cognition. Human problem-solving rarely follows a unidirectional progression; it involves exploration, evaluation of alternatives, backtracking from dead ends, and the synthesis of information from disparate lines of inquiry.

This realization motivated a shift away from the "chain" metaphor toward frameworks based on more complex and flexible data structures: trees and graphs.

3.1. Tree of Thoughts: The Architecture of Exploration

A significant evolution beyond linear constraints was the "Tree of Thoughts" (ToT) framework (Yao et al., 2023). The authors argued that CoT is analogous to an intuitive, stream-ofconsciousness process, whereas complex problems often demand a more deliberate, systematic, and exploratory approach. This distinction explicitly draws upon the dual-process theory in cognitive science, contrasting "System 1" (fast, intuitive) thinking with "System 2" (slow, deliberate) reasoning (Kahneman, 2011). Many tasks require the exploration of different branches of thought and the ability to make strategic judgments—a process poorly modeled by the fixed structure of CoT.

The ToT framework operationalizes this deliberate process by structuring reasoning as a tree, where each node represents a "thought"—a coherent semantic unit serving as an intermediate step. The framework integrates three key components:

1. Thought Generation: At each node, the LLM acts as a "proposer," generating multiple potential next steps, thereby creating distinct branches.

2. State Evaluation: The LLM is then utilized as an "evaluator," assessing the progress and viability of each branch toward reaching a final solution. This evaluation (e.g., a numeric score or classification) provides the necessary feedback for guiding the search.

Table 2. Comparison of Foundational Prompting Strategies for Eliciting Reasoning

Technique	Core Mechanism	Key Strengths	Limitations
Chain-of-Thought (CoT) (Wei et al., 2022b)	Eliciting intermediate reasoning steps via few-shot examples (Input/Reasoning/Output).	Significantly improves multi-step reasoning; high interpretability.	Requires manual creation of task-specific exemplars.
Zero-Shot CoT (Kojima et al., 2022)	Using a generic prompt (e.g., "Let's think step by step") to activate reasoning without examples.	High generality; easy implementation; no manual exemplars needed.	Lower accuracy than Few-Shot CoT; reasoning quality can be inconsistent.
Self-Consistency (Wang et al., 2022)	Generating multiple reasoning paths and selecting the most consistent answer (majority vote).	Robustness to individual reasoning errors; significantly improves accuracy.	High computational cost due to multiple generation passes.
Least-to-Most Prompting (Zhou et al., 2022)	Decomposing a complex problem into sequential subproblems and solving them iteratively.	Handles high compositional complexity; enables generalization.	Requires careful design of Decomposition examples; multi-stage process.

3. Search Algorithm: A systematic search algorithm (e.g., Breadth-First Search (BFS) or Depth-First Search (DFS)) navigates the tree, utilizing the state evaluations to prioritize exploration and prune unpromising branches.

This architecture enables the model to engage in deliberate problem-solving, incorporating lookahead and backtracking capabilities. Experiments on tasks requiring non-trivial planning or search demonstrated substantial performance gains over CoT. For example, in the Game of 24, a task requiring complex combinatorial search, GPT-4 achieved a success rate of only 4% using CoT, whereas the ToT framework boosted performance to 74% (Yao et al., 2023). This marked a critical transition from simple reasoning elicitation to systematic reasoning exploration.

3.2. Graph of Thoughts: Synthesis and Refinement

While ToT expanded reasoning capabilities through branching, Besta et al. (2023) argued for an even more flexible topology. Human cognition often involves the synthesis of information from disparate reasoning paths and the iterative refinement of ideas. Their work, "Graph of Thoughts" (GoT), generalizes the reasoning structure into a directed acyclic graph.

In GoT, thoughts (nodes) and their dependencies (edges) allow for a richer set of cognitive operations than a tree structure permits:

- **Aggregation:** Thoughts from independent branches can be merged and synthesized into a new, more comprehensive thought, allowing the model to combine the strengths of different approaches (e.g., merging two different solution strategies into a superior hybrid).
- **Refinement:** Thoughts can be iteratively improved. The graph structure supports explicit "refinement loops," where a thought can be critiqued and transformed multiple times until it meets a quality standard.

This graph structure inherently captures the complex, non-linear dependencies inherent in elaborate problem-solving. It enables an LRM not only to explore divergent paths but also to weave insights from those paths together, a capability essential for tasks involving complex planning, multi-faceted analysis, or solutions composed of many sub-solutions.

Table 3. Comparison of Reasoning Topologies and Cognitive Analogies in LRMs

Topology	Structure	Key Cognitive Operations	Ideal Use Cases	Cognitive Analogy
Chain (e.g., CoT)	Linear sequence.	Step-by-step decomposition; sequential deduction.	Arithmetic problems; simple logic.	Stream of consciousness (System 1) (Kahneman, 2011).
Tree (e.g., ToT)	Branching structure.	Exploration of alternatives; backtracking; lookahead; systematic search (BFS/DFS).	Complex planning; combinatorial problems (e.g., Game of 24); strategic decisions.	Deliberate exploration; Counterfactual reasoning (System 2) (Kahneman, 2011).
Graph (e.g., GoT)	Directed graph.	Synthesis of disparate paths; aggregation of ideas; iterative refinement loops.	Elaborate problem-solving; multi-faceted analysis; tasks requiring synthesis.	Complex synthesis and integration of knowledge.
Dynamic (e.g., SELFDISCOVER)	Discovered autonomously.	Meta-reasoning; adaptive strategy selection; composition of atomic modules.	Diverse benchmarks; tasks where the optimal strategy is unknown a priori.	Meta-cognition (Reasoning about how to reason).

JAIKE https://doi.org/10.37819/Jaike.25.0010

3.3. Meta-Reasoning: The SELF-DISCOVER Framework

A crucial insight advanced by Zhou et al. (2024) is that the optimal reasoning structure is inherently task-dependent; no single pre-defined structure is universally optimal. Instead of imposing a structure (chain, tree, or graph), their SELF-DISCOVER framework empowers the LLM to autonomously formulate a bespoke reasoning plan.

The framework operates by allowing the model to select relevant atomic reasoning modules (e.g., "Critical Thinking," "Decomposition," "Analogy Generation") and then structure these modules into an explicit reasoning plan tailored to the specific problem. This meta-reasoning capability—reasoning about *how* to reason—allows the model to dynamically adapt its problem-solving strategy on the fly. This self-discovered approach has shown improved performance on challenging benchmarks such as BIG-Bench Hard (Suzgun et al., 2022), demonstrating a trajectory toward more flexible, efficient, and autonomous reasoning systems. A comparison of these advanced reasoning topologies is provided in Table 3.

3.4. Synergies and Interconnections

It is crucial to recognize that these advanced reasoning structures (Pillar II) do not operate in isolation but exhibit strong synergies with the other pillars. The effectiveness of exploratory frameworks like ToT and GoT often depends on the quality of the underlying reasoning elicited (Pillar I). Furthermore, the "State Evaluation" step in ToT frequently relies on the LRM's ability to utilize external tools (Pillar III) for grounding or verification, or to apply self-critique mechanisms (Pillar IV) to assess the viability of a reasoning path. This interdependence (visualized in Figure 1) highlights the cohesive nature of the LRM framework, where advancements in one area potentiate the capabilities of the others.

3.5. The Accuracy-Efficiency Trade-off

While the evolution from linear chains to complex graphs significantly enhances reasoning accuracy and robustness, this progression introduces substantial computational overhead. The different methodologies present distinct trade-offs between accuracy, latency, and computational cost.

Chain-of-Thought (CoT) is computationally efficient, typically requiring only a single forward pass (inference call) from the LLM. However, its accuracy is limited by the fragility of that single reasoning path. Self-Consistency increases robustness by sampling multiple paths, linearly increasing the computational cost (e.g., 10 samples equate to 10x the cost of CoT).

Frameworks like ToT and GoT introduce a multiplicative increase in cost. These methods require numerous LLM calls not only for generating potential thoughts (proposing branches) but also for evaluating the state of those thoughts to guide the search. In complex tasks, the search space can become vast, leading to hundreds or thousands of inference calls for a single problem. This high computational demand and increased latency currently limit the practical deployment of these advanced methods in real-time applications, highlighting the critical need for research into more efficient reasoning architectures (as discussed in Section 6).

4. THE GROUNDING PROBLEM: TOOL-AUGMENTED AND PROGRAM-AIDED REASONING

A fundamental limitation of purely neural LLMs is their inherent detachment from the external world. An LLM's knowledge is static (frozen at the end of training), its mathematical capabilities are approximate rather than precise, and it lacks the ability to interact with real-time systems.

This leads to factual inaccuracies ("hallucinations"), calculation errors, and an inability to utilize up-to-date information. The solution to this "grounding problem" has been to augment LLMs with the capacity to use external tools and interact with formal systems, transforming them from isolated knowledge repositories into interactive, grounded agents.

4.1. Offloading Computation: Program-Aided Language Models

Despite their linguistic fluency, LLMs are inherently poor at precise arithmetic and symbolic manipulation, as their processing relies on pattern matching rather than formal computation. To address this deficiency, Gao et al. (2022) introduced Program-Aided Language Models (PAL).

The core innovation is to reframe the task: instead of prompting the LLM to generate a final numerical answer, it is prompted to generate a program (e.g., in Python) that, when executed by a deterministic interpreter, yields the solution.

This approach leverages the complementary strengths of the two systems: the LLM excels at natural language understanding and decomposing the problem into logical, programmatic steps, while the external interpreter executes the computation with perfect fidelity. This synergy dramatically improves accuracy in mathematical reasoning and enhances interpretability.

4.2. Formal and Quantitative Reasoning

While PAL addresses computation via code, specialized efforts have focused on enhancing reasoning in formal and quantitative domains. Minerva (Lewkowycz et al., 2022), fine-tuned on extensive mathematical and scientific corpora, demonstrated significant improvements in solving complex quantitative problems by learning to utilize appropriate mathematical notation and concepts, even without relying on external execution.

Furthermore, the pursuit of truly verifiable reasoning has led to the integration of LLMs with formal verification systems and theorem provers. Frameworks such as LeanDojo (Yang et al., 2023) enable LLMs to interact with proof assistants (e.g., Lean). In this context, the theorem prover acts as an external environment that provides rigorous, deterministic feedback on the logical validity of the model's proposed steps. This grounding in formal logic addresses the critical challenge of faithfulness, ensuring that the reasoning is not merely plausible but mathematically sound. Despite these advancements, significant limitations remain in formal reasoning. A primary hurdle is *autoformalization*—the process of accurately translating natural language problems into formal logical representations. Current models often struggle with the ambiguity and implicit context of natural language, leading to brittle translations that derail the subsequent proving process. Additionally, theorem proving itself faces the challenge of navigating a vast, combinatorial search space. While LLMs can suggest plausible steps, the search for a complete proof remains computationally intractable for complex theorems, often requiring sophisticated heuristics and human expertise that LRMs have yet to fully replicate.

4.3. Dynamic Interaction: Toolformer and ReAct

The concept of tool use was generalized and scaled in "Toolformer" (Schick et al., 2023). This work introduced a self-supervised method for teaching an LLM to utilize external APIs (e.g., search engines, calculators, translation systems) without requiring massive human-annotated datasets. The model learns to autonomously decide when to call a tool, what arguments to pass, and how to integrate the tool's output back into its generation process.

Complementing this, the ReAct (Reason and Act) framework (Yao et al., 2022) established a powerful paradigm for interleaving reasoning and action. In ReAct, the model generates alternating "thought" (internal reasoning about the plan and task updates) and "action" (interaction with an external tool, such as a knowledge base lookup) steps. The result of the action (an "observation") is fed back into the context, prompting the next cycle. This tight feedback loop enables the model to construct dynamic, adaptable plans, closely mirroring human problem-solving in information-rich environments.

4.4. The LRM as an Orchestrator and Agent

The principles underlying PAL, Toolformer, and ReAct reflect a broader conceptual shift: positioning the LRM not as a monolithic problem-solver, but as the "reasoning engine" at the core of a larger ecosystem. This is evident in the design of contemporary flagship models like GPT-4 (OpenAI, 2023) and Gemini (Hassabis and the Gemini team, 2023).

This integration of reasoning and action is foundational to the rapidly growing field of Agentic AI. The LRM serves as the cognitive core for systems capable of executing complex, long-horizon tasks. This burgeoning field includes frameworks like AutoGPT and Voyager, as well as research into complex behavioral simulations, such as Generative Agents (Park et al., 2023), which utilize LRMs for planning, memory management, and reflection.

This paradigm is further supported by techniques such as Retrieval-Augmented Generation (RAG) (Lewis et al., 2020), where the LLM's context is dynamically supplemented with information retrieved from external knowledge corpora. A retriever module first fetches relevant documents, which are provided to the LLM along with the prompt. This allows the model to ground its responses in specific, verifiable, and up-to-date information, significantly mitigating hallucination. Collectively, these approaches, summarized in Table 4, delineate a future where the LRM functions as a central orchestrator and autonomous agent, intelligently delegating sub-tasks to a suite of specialized, reliable external tools.

JAIKE https://doi.org/10.37819/Jaike.25.0010

Table 4. Key Frameworks and Strategies for Tool-Augmented and Grounded Reasoning

Framework/ Strategy	Core Mechanism	Primary Problem Addressed	Examples of Tools
PAL (Program-Aided) (Gao et al., 2022)	LLM generates code; external interpreter executes the code to find the solution.	Inaccurate arithmetic and symbolic manipulation.	Python Interpreter; Symbolic Solvers.
Formal Verification (e.g., Lean-Dojo (Yang et al., 2023))	LLM interacts with formal theorem provers to generate and verify proofs.	Lack of formal verifiability and faithfulness.	Theorem Provers (e.g., Lean).
Toolformer (Schick et al., 2023)	Training LLMs via selfsupervision to generate API calls and incorporate the results.	Limitations in specialized tasks and access to realtime information.	Calculators; Search Engines; Translation Systems.
ReAct (Reason+Act) (Yao et al., 2022) / Agentic AI	Interleaving internal reasoning ("thoughts") with external actions ("actions") in a tight feedback loop, enabling autonomous goal execution.	Inability to dynamically adjust plans based on external feedback; lack of agency.	Knowledge Bases; Web APIs; Interactive Environments.
RAG (Retrieval-Augmented) (Lewis et al., 2020)	Dynamically retrieving relevant documents and injecting them into the LLM's context before generation.	Static knowledge cutoffs; factual hallucination.	Vector Databases; Document Stores; Search Indices.

5. The Frontier of Reliability: Evaluation, Critique, and Self-Correction

As the capabilities of LRMs expand, the imperative for rigorous evaluation and the challenge of ensuring reliability have become central concerns. This research frontier focuses on moving beyond simple accuracy metrics to understand the fundamental nature of machine reasoning and developing mechanisms for autonomous self-improvement—critical prerequisites for deployment in high-stakes domains.

5.1. The Measurement Problem: Defining "Reasoning"

Evaluating the reasoning capabilities of LRMs remains a profound challenge. Standard benchmarks can sometimes be exploited by models that learn superficial statistical patterns without engaging in genuine, generalized reasoning. This has led to the development of more robust evaluation suites like MMLU (Hendrycks et al., 2020) and BIG-Bench (Suzgun et al., 2022).

Furthermore, comprehensive frameworks such as the Holistic Evaluation of Language Models (HELM) (Liang et al., 2022) emphasize the necessity of evaluating models across a broad spectrum of scenarios, metrics (including fairness and robustness), and potential biases.

A fundamental epistemological question persists: are these models reasoning from first principles, or executing a highly sophisticated form of pattern matching? Research from Microsoft (Research, 2023) explores this by systematically altering problem parameters to test a model's adaptive reasoning capacity. Findings often indicate that while models excel at familiar patterns, performance degrades significantly as problems deviate from the training distribution. This suggests that robust, abstract reasoning remains an area requiring substantial progress.

5.2. The Challenge of Autonomous Self-Correction

A hallmark of advanced intelligence is the ability to recognize and rectify one's own errors. A critical investigation by Huang et al. (2023), Large Language Models Cannot Self-Correct Reasoning Yet," provided a sobering assessment of this capability.

The study found that when an LLM is prompted to re-evaluate its own incorrect answer without external feedback, it frequently fails to identify its mistakes. In many cases, models express low confidence but arrive at the same incorrect conclusion, often reinforcing the flawed reasoning. This indicates that genuine, reliable self-correction is not an inherent emergent property of scale and necessitates dedicated mechanisms and training paradigms.

5.3. Scaffolding Self-Improvement

In response to this challenge, frameworks have been proposed to provide structured processes for iterative improvement.

These approaches generally fall into two categories: iterative refinement based on generated feedback, and principle-based supervision.

Iterative Refinement: "Self-Refine" (Madaan et al., 2023) establishes a structured, iterative loop that decomposes the refinement process into discrete steps. First, the model generates an output. Second, the model (often the same instance) is prompted to act as a critic, generating constructive feedback on its own output. Finally, the model uses this specific feedback to refine the initial solution. This cycle of "generate, provide feedback, refine" can be repeated. While this process remains externally scaffolded—requiring a designed prompting loop rather than intrinsic self-awareness—it demonstrates empirical improvements in output quality across various domains by forcing the model to consider its own deficiencies. It represents a concrete step toward developing the self-critical capabilities necessary for autonomous improvement, though its efficacy is contingent on the model's ability to generate accurate critique.

Principle-Based Supervision: A distinct and more formalized approach, focused primarily on safety and alignment, is Constitutional AI (CAI) (Bai et al., 2022). Instead of relying solely on human feedback (Reinforcement Learning from Human Feedback - RLHF), which can be inconsistent and difficult to scale, CAI uses the AI model itself for supervision (Reinforcement Learning from AI Feedback - RLAIF). The process involves training a model to critique and revise its own responses based on an explicit set of guiding principles (a "constitution").

This allows for the internalization of desired behaviors and norms in a transparent manner.

While developed for alignment (e.g., ensuring harmlessness), the underlying principle of using a formalized framework to guide self-correction offers a promising path toward making reasoning refinement more reliable and interpretable than ad-hoc refinement. These approaches are compared in Table 5.

6. DISCUSSION AND FUTURE DIRECTIONS

The evolution of Large Reasoning Models-from simple prompting interventions to complex, self-discovering, tool-augmented frameworks-represents a remarkable trajectory in artificial intelligence designation refers to this entire augmented system research. It is crucial to reiterate that the LRM

Table 5. Comparison of Approaches to Self-Correction and Refinement in LRMs

Approach	Feedback Source	Mechanism	Primary Goal/ Observation
Unguided Self-Correction	Internal (Model itself).	Prompting the model to check its work without specific guidance or external input.	Found to be largely ineffective; models often reinforce errors (Huang et al., 2023).
Scaffolded Iterative Refinement (e.g., Self-Refine (Madaan et al., 2023))	Internal (Model acts as critic).	Structured prompting loop: Generate → Provide Feedback → Refine. Requires external initiation.	Improving the quality and accuracy of the generated output through iteration.
Principle-Based Critique (e.g., Constitutional AI (Bai et al., 2022))	Externalized Principles.	Training the model to critique and revise outputs based on a formalized "constitution."	Aligning model behavior with desired norms (safety, reliability) using AI supervision.

https://doi.org/10.37819/Jaike.25.0010

(LLM plus scaffolding), rather than a modification of the underlying LLM architecture itself. The dominant narrative is one of scaffolding: the research community has progressively constructed a sophisticated set of external and internal supports to guide the latent capabilities of scaled LLMs, transforming them from unpredictable pattern generators into more deliberate and verifiable problem-solvers. The four pillars reviewed form a cohesive intellectual arc that systematically addresses the core limitations of the underlying technology.

However, this rapid innovation has highlighted several profound challenges that will define the next phase of LRM research and are critical to the responsible deployment of this technology.

These challenges are summarized in Table 6.

Faithful and Verifiable Reasoning: The interpretability offered by frameworks like CoT is valuable, but incomplete. The reasoning steps generated by a model are still linguistic artifacts, and there is no formal guarantee that they are logically sound or causally linked to the final output. A model may produce a plausible-sounding justification that supports a hallucinated conclusion. This lack of faithfulness poses significant risks in high-stakes domains. For instance, in clinical diagnosis, an LRM might generate a compelling explanation for an incorrect diagnosis based on misinterpreted data; in legal interpretation, a model might construct a plausible argument supporting a flawed legal conclusion. Developing methods to ensure that reasoning is not merely plausible but *faithful* and verifiably correct is a central open problem. Research into grounding LRMs in formal systems, such as interactive theorem provers (Yang et al., 2023), offers a promising direction for achieving rigorous verification.

Computational Efficiency and Scalability: The most advanced reasoning techniques, particularly those involving exploration (ToT, GoT) and ensemble methods (Self-Consistency), incur significant computational costs, as detailed in Section 3.5. Generating and evaluating numerous reasoning paths for a single query is impractical for many real-world applications.

This highlights a fundamental trade-off between reasoning quality and computational efficiency.

While deliberate, System 2-like exploration significantly improves accuracy, the associated inference cost presents a substantial barrier. It remains an open question whether this cost is an inherent limitation of applying deliberate reasoning to current architectures, or a temporary engineering hurdle. Potential solutions include the development of learned heuristics to aggressively prune the search space, the distillation of complex reasoning paths into more efficient models, or the development of novel architectures intrinsically capable of efficient deliberation.

Sustainability and Environmental Impact: The reliance on massive scale for emergent reasoning abilities, coupled with the high computational demands of advanced reasoning architectures during inference, raises significant concerns regarding the environmental impact of LRMs. The energy consumption and associated carbon footprint of training and deploying these models are substantial (Strubell et al., 2019). As the field moves towards increasingly complex scaffolding techniques, prioritizing computational sustainability and developing energy-efficient reasoning methods becomes imperative for responsible innovation.

Autonomous Learning and Correction: Models currently lack the intrinsic ability to reliably detect their own failures (Huang et al., 2023). While scaffolded methods like Self-Refine (Madaan et al., 2023) are promising, the long-term objective is the development of models capable of learning from mistakes autonomously and in an online fashion. Moving beyond prompted refinement to create systems that intrinsically identify and correct flawed reasoning without external loops remains a significant challenge toward more general intelligence.

Theoretical Understanding: The mechanisms by which reasoning emerges in these models, particularly as a function of scale, remain poorly understood. What are the precise neural mechanisms underpinning these complex behaviors? A deeper, theoretical grounding of these emergent abilities, potentially drawing from cognitive science (Kahneman, 2011) and theoretical computer science, is essential for building predictable, robust, and safe systems.

The Ethics of Autonomous Reasoners: As LRMs become increasingly capable of autonomous planning, tool use, and self-correction, they begin to function as autonomous agents (Park et al., 2023). This raises profound ethical questions concerning accountability, control, and value alignment.

Ensuring that tool-augmented LRMs operate within safe boundaries and adhere to robust ethical principles (such as those explored in Constitutional AI (Bai et al., 2022)) in novel situations requires a sustained, multi-disciplinary effort.

7. CONCLUSION

The field of Large Reasoning Models has successfully transitioned from an initial phase of discovering the emergent cognitive abilities of scaled LLMs to a systematic engineering discipline focused on harnessing those abilities. The central thesis of this review—that the advancement of machine reasoning is fundamentally driven by the systematic construction of cognitive scaffolding around foundational LLMs-is strongly supported by the literature synthesized.

The progression from the linear Chain-of-Thought to the exploratory Tree of Thoughts, the synthetic Graph of Thoughts, and the adaptive SELF-DISCOVER framework marks a clear trajectory in the complexity and power of these scaffolds. Concurrently, the integration of external tools, formal systems, and agentic frameworks has been pivotal in grounding these models, tethering their abstract reasoning capabilities to the verifiable, external world, and enabling the development of increasingly autonomous agents.

We have moved beyond viewing LLMs as mere text generators; they are now understood as latent reasoning engines whose potential is unlocked through this carefully constructed scaffolding, forming the functional systems we term LRMs. The techniques reviewed herein represent the foundational components of this emerging science of machine cognition. Yet, the path for-ward is defined

Table 6. Open Challenges and Future Research Directions in LRM Development

Challenge Area	Core Problem	Potential Research Avenues
Faithfulness and Verification	Generated reasoning steps may be plausible but logically unsound or non-causal, posing risks in highstakes domains (e.g., medicine, law).	Formal verification methods; Integration with theorem provers; Causal reasoning frameworks; Enhanced mechanistic interpretability.
Efficiency and Scalability	Advanced techniques (e.g., ToT, Self-Consistency) are computationally expensive, limiting real-world application.	Efficient search algorithms; Learned evaluation heuristics; Knowledge distillation for reasoning processes; Novel architectures for deliberation.
Sustainability and Environment	High energy consumption and carbon footprint associated with training and inference of scaled LRMs.	Energy-efficient algorithms; Hardware optimization; Computational sustainability analysis (Strubell et al., 2019).
Autonomous Correction	Models struggle to reliably identify and correct their own errors without external feedback loops.	Development of intrinsic selfevaluation mechanisms; Online learning from errors; Meta-learning for correction strategies.
Theoretical Understanding	The mechanisms by which reasoning emerges with scale are poorly understood, hindering predictability.	Deeper connections with cognitive science; Formal theories of emergent cognition; Rigorous theoretical grounding.
Ethics and Alignment	Increased autonomy through tool use and self-correction raises risks of misalignment and harmful actions.	Robust adversarial testing; Transparent and controllable value alignment frameworks (e.g., Constitutional AI); Safety protocols for agents.

JAIKE https://doi.org/10.37819/Jaike.25.0010

by the critical challenges of reliability, verifiability, and genuine self-correction.

Addressing these challenges is the prerequisite for deploying these powerful technologies safely and ethically, realizing a future where AI systems serve not just as information processors, but as trustworthy and capable partners in human problem-solving.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

There is no requirement for ethical approval and consent of participants.

Human and Animal Rights

No humans and animals are involved in this designed study.

Consent for Publication

There is no requirement of consent for personal details, audio-video material, etc.

Availability of Data and Materials

The sources of data and materials have been mentioned in the manuscript in support of the findings.

Funding

The author declare that no funding sources are involved in this manuscript.

Acknowledgements

All author are thankful to their parent institutions.

REFERENCES

Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, et al. Solving quantitative reasoning problems with language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training, 2018.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, et al. Self-refine: Iterative refinement with self-feedback, 2023.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. Measuring massive multitask language understanding, 2020.

Daniel Kahneman. Thinking, Fast and Slow. *Farrar, Straus and Giroux*, 2011.

Demis Hassabis and the Gemini team. Introducing gemini: our largest and most capable ai model. *Google AI Blog*, 2023.

Denny Zhou, Nathanael Sch.rli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, et al. Leastto-most prompting enables complex reasoning in large language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 2021.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022b.

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, et al. Emergent abilities of large language models. *Transactions on Machine Learning Research*, 2022a.

Jie Huang, Xinyun Chen, Swaroop Mishra, Shibo Zheng, Adams Wei Yu, Xinying Song, and Denny Zhou. Large language models cannot self-correct reasoning yet, 2023.

Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST)*, 2023.

Kaiyu Yang, Aidan Swope, Alex Gu, Rahul Chisholm, Nikhil Murthy, Peiyang L.pez-Guevara, Thorsten Rabe, Anand Bansal, and Santiago Labella. Leandojo: Theorem proving with retrieval-augmented language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Luyu Gao, Aman Madaan, Shuyan Zhou, Uri Alon, Pengfei Liu, Yiming Yang, Graham Neubig, and Karthik Narasimhan. Program-aided language models, 2022.

Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Lukas Gianinazzi, Joanna Gajda, et al. Graph of thoughts: Solving elaborate problems with large language models, 2023.

Microsoft Research. A ladder of reasoning: Testing the power of imagination in llms. *Microsoft Research Blog*, 2023.

Mirac Suzgun, Nathan Scales, Nathanael Sch.rli, Sebastian Gehrmann, Yi Tay, Hyung Won Chung, et al. Challenging big-bench tasks and whether chain-of-thought can solve them, 2022.

OpenAI. *Gpt-4 technical report*, 2023.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Pei Zhou, Jay Pujara, Xiang Ren, Swaroop Mishra, Shibo Zheng, Denny Zhou, et al. Large language models self-discover reasoning structures, 2024.

Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, et al. Holistic evaluation of language models. *Transactions on Machine Learning Research (TMLR)*, 2022.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L Griffiths, Yue Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yue Cao. React: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Timo Schick, Jane Dwivedi-Yu, Roberto Dess., Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models that teach themselves to use tools. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

XuezhiWang, JasonWei, Dale Schuurmans, Quoc Le, Ed Chi, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, et al. Constitutional ai: Harmlessness from ai feedback, 2022.