Academic Journal of Research and Scientific Publishing | Vol 8 | Issue 88
Publication Date: 5 August 2026 ISSN: 2706-6495

An Analytical Framework for the Evolution of Large Language Models: From Reasoning to LLM-as-a-Judge

Mariam Khaled Obeidat

PhD in Computer Science, King Hussein School of Computing Sciences, Princess Sumaya University for Technology, Hashemite Kingdom of Jordan

Email: mariam.kh.obeidat@hotmail.com | ORCID: 0009-0007-6603-5176

Received:
7 May 2026
First Decision:
14 May 2026
Revised:
20 June 2026
Accepted:
14 July 2026
Published:
5 August 2026

Abstract:

This study aims to develop an integrative computational-analytical framework for explaining the functional evolution of large language models through four interrelated dimensions: Reasoning, Retrieval and Knowledge Grounding, Alignment, and Judging, collectively represented by the RRAJ framework. The study adopts a systematic analytical literature review of research published between 2020 and 2026, drawing on Scopus, Web of Science, and IEEE Xplore, and applying comparative and thematic analysis to trace major technical transformations and functional relationships across the four dimensions. The findings indicate that the evolution of large language models is no longer driven primarily by model scaling, but increasingly by system scaling, in which knowledge, computation, and control are distributed across model parameters, contextual information, external knowledge sources, inference-time computation, and evaluation mechanisms. The analysis further reveals a shift from isolated capabilities toward functional convergence among RRAJ components, with LLM-as-a-Judge emerging as an important transition that enables evaluation to support diagnosis, feedback, and correction. The study concludes that future LLM systems are likely to become more adaptive, verifiable, self-evaluating, and self-correcting. It further recommends improving judge reliability, error traceability, and human oversight in high-risk applications, while exploring judge-governed execution as a future extension for agentic and robotic systems.

Keywords: Large language models; RRAJ; reasoning; retrieval-augmented generation; alignment; LLM-as-a-Judge; functional convergence; self-evaluation; self-correction.

1. Introduction:

In recent years, the field of natural language processing (NLP) has undergone a rapid transformation with the emergence of large language models (LLMs), which have evolved from specialized models primarily designed for text prediction into general-purpose systems capable of performing a broad range of linguistic and cognitive tasks. The introduction of GPT-3 marked a significant milestone in this transformation, demonstrating that scaling model size and pretraining on large-scale datasets can enable models to perform previously unseen tasks through in-context learning, including zero-shot and few-shot learning, without requiring task-specific retraining (Brown et al., 2020). This development contributed to establishing the paradigm of general-purpose language models capable of adapting to multiple tasks through natural-language instructions and textual context.

However, subsequent applications have demonstrated that the ability to generate convincing text or adapt from a limited number of examples does not necessarily imply that a model possesses reliable reasoning capabilities, access to up-to-date knowledge, or the ability to align with user preferences. Consequently, research efforts have increasingly focused on enhancing the reasoning capabilities of LLMs. Wei et al. (2022) showed that Chain-of-Thought (CoT) prompting, which encourages a model to generate intermediate reasoning steps before producing a final answer, can improve performance across a range of arithmetic and logical reasoning tasks. Similarly, Kojima et al. (2022) demonstrated that some of these reasoning capabilities can be elicited in a zero-shot setting without providing prior reasoning exemplars. This line of research subsequently evolved toward more sophisticated strategies involving the exploration of multiple solution paths and the evaluation of intermediate reasoning steps, as exemplified by the Tree of Thoughts framework (Yao et al., 2023). Reasoning has thus emerged as a fundamental dimension in understanding the evolution of the cognitive capabilities of large language models.

Alongside reasoning, another major challenge has emerged concerning the limitations of the knowledge encoded in a model's parameters, including the possibility that such knowledge may become outdated or that its underlying sources may be difficult to verify. In this context, Lewis et al. (2020) introduced the concept of Retrieval-Augmented Generation (RAG), which combines a model's parametric knowledge with external knowledge sources that can be retrieved during the generation process. This approach enables models to ground their outputs in updatable external information rather than relying exclusively on knowledge acquired during pretraining. The concept was subsequently extended toward more integrated architectures, such as Self-RAG, which combines retrieval, generation, and self-reflection to assess the relevance and quality of retrieved evidence and generated outputs (Asai et al., 2024). This progression suggests that improving the performance of LLMs

depends not only on stronger reasoning capabilities, but also on the quality, relevance, and reliability of the knowledge and evidence used during response generation.

In parallel, increasing attention has been directed toward aligning model behavior with users' instructions and preferences. Large-scale pretraining alone does not necessarily ensure that a model will generate responses that are useful or consistent with human intent. Ouyang et al. (2022) demonstrated that Reinforcement Learning from Human Feedback (RLHF) can improve a model's ability to follow instructions and produce outputs that are better aligned with human preferences. Chung et al. (2024) further showed that scaling instruction fine-tuning and incorporating reasoning-related data into the fine-tuning process can enhance generalization to previously unseen tasks. Subsequently, Rafailov et al. (2023) introduced Direct Preference Optimization (DPO), a method that enables models to be optimized directly from human preference data without requiring the full conventional RLHF pipeline. Collectively, these developments indicate that progress in LLMs is increasingly determined not only by a model's ability to generate an answer, but also by how effectively its behavior can be steered toward desired responses.

As model outputs have become increasingly complex, evaluating response quality has emerged as a research challenge in its own right. Traditional metrics based on matching generated outputs against reference answers are not always well suited to evaluating open-ended responses for which multiple valid answers may exist. Consequently, large language models themselves have increasingly been employed as evaluators. Liu et al. (2023) introduced G-Eval, a framework that uses a large language model together with predefined evaluation criteria to assess the quality of generated text. Zheng et al. (2023), meanwhile, examined the broader concept of LLM-as-a-Judge through MT-Bench and Chatbot Arena, demonstrating that LLMs can be used to evaluate and compare model responses while achieving high levels of agreement with human preferences. However, they also identified several sources of bias, including position bias, verbosity bias, and self-enhancement bias. These findings indicate that replacing human evaluation with LLM-based evaluation does not eliminate concerns regarding reliability; rather, it reformulates them in a different form. More recent studies further suggest that reliability, consistency, and bias have become central issues in the development of LLM-as-a-Judge frameworks (Gu et al., 2026).

Taken together, these trends suggest that the evolution of large language models cannot be understood merely as an increase in model scale or generative capacity. Rather, it reflects a broader transformation in their functional roles: from in-context learning and text generation to reasoning, the integration of external knowledge, alignment with human preferences, and, ultimately, the evaluation of outputs produced by artificial intelligence systems.

Moreover, recent developments, such as Self-RAG and the incorporation of Chain-of-Thought reasoning into fine-tuning and evaluation processes, indicate that these capabilities are becoming increasingly intertwined rather than evolving along separate and independent trajectories (Asai et al., 2024; Chung et al., 2024; Liu et al., 2023).

Although substantial bodies of literature have emerged around reasoning, retrieval augmentation, alignment, and LLM-as-a-Judge, there remains a need for an analytical framework that integrates these research directions and explains the functional relationships among them. Accordingly, the present study proposes the RRAJ (Reasoning–Retrieval–Alignment–Judging) framework to analyze the evolution of large language models from an integrative functional perspective. The framework does not assume that these dimensions emerged in a strict chronological sequence; rather, it conceptualizes them as overlapping capabilities that have developed partly in parallel and have increasingly converged within modern LLM-based systems. On this basis, the study aims to analyze the literature published between 2020 and 2026, identify the major shifts and interrelationships among these four dimensions, and ultimately develop a framework that explains the growing movement toward large language models that integrate reasoning, knowledge grounding, alignment, and evaluation.

1.1. Research Problem

The research problem addressed in this study arises from the increasing fragmentation of the literature on large language models across specialized research streams focusing on reasoning, retrieval augmentation, preference alignment, and LLM-based evaluation. At the same time, contemporary LLM-based systems increasingly integrate multiple such functions within a single architecture or workflow.

Examining each research stream in isolation makes it difficult to explain the broader transformation in the nature and functional role of LLMs: how have they evolved from models primarily designed for language generation and adaptation to textual examples into systems capable of reasoning over problems, retrieving supporting evidence, adapting to user preferences, and ultimately participating in the evaluation of AI-generated outputs?

Accordingly, the central research problem lies in the need for an integrative analytical framework that explains the relationships among these transformations and clarifies how they collectively contribute to reshaping the functional role of large language models.

1.2. Research Gap

The research gap does not stem from a lack of studies on reasoning, Retrieval-Augmented Generation (RAG), alignment, or LLM-as-a-Judge, as each of these areas has already developed a substantial body of literature and a growing number of specialized reviews.

Rather, the gap addressed by the present study lies in the limited availability of analytical frameworks that examine reasoning, retrieval, alignment, and evaluation as interconnected components of a broader functional evolution in large language models, and that analyze the emerging relationships and interactions among these capabilities rather than treating them as separate research trajectories.

Accordingly, this study focuses on the concept of Functional Convergence, referring to the increasing integration and interdependence of these capabilities, as a key characteristic of the evolution of modern LLMs.

1.3. Aim of the Study

This study aims to develop an integrative computational-analytical framework for explaining the functional evolution of large language models by examining the development, interaction, and convergence of four key dimensions: Reasoning, Retrieval and Knowledge Grounding, Alignment, and Judging (RRAJ).

Specific Objectives

The study seeks to:

1. Analyze the major transformations in reasoning mechanisms within large language models, from in-context learning to multi-step and extended inference-time reasoning.
2. Examine how retrieval and knowledge-grounding techniques have extended LLM capabilities beyond knowledge encoded in model parameters through access to external sources.
3. Analyze the evolution of alignment mechanisms from instruction tuning and Reinforcement Learning from Human Feedback (RLHF) toward direct preference optimization approaches.
4. Trace the development of LLM-as-a-Judge as an emerging evaluation paradigm and examine its role, reliability, and major sources of bias.
5. Compare the four RRAJ dimensions according to their computational mechanisms, locations of intervention, sources of knowledge or feedback, and stages of execution.
6. Analyze the functional relationships and convergence among Reasoning, Retrieval, Alignment, and Judging, particularly in systems that integrate two or more of these capabilities.
7. Identify the major open challenges and future research directions associated with the evolution of RRAJ toward increasingly adaptive, self-evaluating, and self-correcting AI systems.

1.4. Research Questions

1. RQ1: How have reasoning mechanisms in large language models evolved from in-context learning toward multi-step and extended inference-time reasoning?

2. RQ2: How have retrieval and knowledge-grounding techniques changed the dependence of large language models on knowledge encoded in their parameters?
3. RQ3: How have alignment mechanisms evolved from instruction tuning and Reinforcement Learning from Human Feedback (RLHF) toward direct preference optimization approaches?
4. RQ4: How has LLM-based evaluation evolved into the LLM-as-a-Judge paradigm, and what role does Judging play within the broader LLM lifecycle?
5. RQ5: What computational mechanisms and levels of intervention distinguish the four dimensions of the RRAJ framework?
6. RQ6: How do Reasoning, Retrieval, Alignment, and Judging interact, and to what extent does the literature indicate a transition from separate capabilities toward functional convergence?
7. RQ7: What major challenges and future research directions emerge from the RRAJ framework, particularly toward increasingly adaptive, self-evaluating, and self-correcting AI systems?

2. Conceptual Background and Analytical Framework:

2.1. Functional Evolution of Large Language Models:

This study is grounded in the concept of the functional evolution of large language models, which suggests that the development of these models should not be understood solely in terms of increases in model scale or parameter count, but also in terms of changes in the nature of the functions they are capable of performing and the sources of knowledge and signals on which they rely when carrying out tasks. Early large-scale models demonstrated important capabilities in in-context learning and in performing novel tasks from only a limited number of examples (Brown et al., 2020). Subsequent developments, however, have moved toward more complex capabilities, including multi-step reasoning, the use of external knowledge, behavioral alignment with human preferences, and the evaluation of output quality. This perspective is further supported by findings showing that performance improvements do not necessarily depend on increasing model size alone. Ouyang et al. (2022), for example, demonstrated that smaller models, once aligned using human feedback, could be preferred over larger unaligned models under the evaluation conditions considered in their study. Accordingly, the present study conceptualizes the evolution of LLMs as a transition from general-purpose generative capability toward a set of interconnected cognitive, behavioral, and evaluative capabilities.

2.2. The Four Dimensions of the RRAJ Framework:

This study proposes an analytical framework denoted as RRAJ: Reasoning–Retrieval–Alignment–Judging, comprising four core dimensions that are used to classify the literature and analyze the functional evolution of large language models.

1- Reasoning represents the dimension concerned with a model's ability to address problems that require intermediate steps rather than directly mapping an input to a final answer. Chain-of-Thought prompting highlighted the importance of generating intermediate reasoning steps for improving model performance across a range of complex tasks (Wei et al., 2022). The literature subsequently expanded toward strategies involving multiple solution paths, search, and iterative evaluation, as exemplified by Tree of Thoughts (Yao et al., 2023).

2- Retrieval and Knowledge Grounding concerns the source of knowledge used during response generation. Retrieval-Augmented Generation (RAG) demonstrated the possibility of combining knowledge encoded in model parameters with external information that can be retrieved during generation (Lewis et al., 2020). This approach later evolved toward more adaptive systems capable of determining when retrieval is needed and assessing the relevance and usefulness of retrieved evidence, as illustrated by Self-RAG (Asai et al., 2024).

3- Alignment represents the behavioral dimension of the framework and concerns the extent to which model responses conform to user instructions, preferences, and desired behavioral criteria. Instruction tuning and Reinforcement Learning from Human Feedback (RLHF) emerged as key mechanisms for shifting models from general text generation toward more effective adherence to human instructions (Ouyang et al., 2022). This line of development later extended to more direct approaches for learning from preference data, such as Direct Preference Optimization (DPO), which enables model optimization from preference pairs without requiring the full conventional pipeline of reward modeling and reinforcement learning (Rafailov et al., 2023).

4- LLM-as-a-Judge, refers to the transition of large language models from serving solely as response generators to participating in the evaluation, ranking, or comparison of outputs produced by other models. Research on G-Eval and MT-Bench has demonstrated the feasibility of using LLMs to assess output quality with varying degrees of agreement with human evaluation (Liu et al., 2023; Zheng et al., 2023). However, this use introduces important challenges, including position bias, verbosity bias, self-enhancement bias, and concerns regarding judgment reliability, making model-based evaluation itself a domain that requires systematic analysis and calibration.

2.3. Operationalization of the Framework in the Analysis:

In this study, the RRAJ framework is employed as an analytical tool for classifying and comparing the literature according to a unified set of dimensions. These dimensions include the primary function of each technique, the source of knowledge or signals on which it relies, the role of human input, the mechanism through which decisions or outputs are produced, the criteria used to assess success, and the principal sources of error or bias.

In simplified terms, Reasoning represents a cognitive capability, whereas Retrieval represents an epistemic capability grounded in external knowledge and evidence. Alignment represents a behavioral capability, while LLM-as-a-Judge represents an evaluative capability. This categorization provides a common analytical basis for examining both the distinct functions of the four dimensions and the ways in which they increasingly interact within contemporary large language model systems.

3. Research Methodology:

This study adopts a Systematic Analytical Literature Review approach to trace and analyze the functional evolution of large language models across four principal dimensions: Reasoning, Retrieval and Knowledge Grounding, Alignment and Preference Optimization, and LLM-as-a-Judge-based Evaluation.

3.1. Search Strategy and Data Sources:

The literature search covered studies published in English between 2020 and August 2026. The primary bibliographic databases used were Scopus, Web of Science Core Collection, and IEEE Xplore. Publisher websites and official conference proceedings were also consulted to verify publication details and identify the final published versions of the included studies. The year 2020 was selected as the starting point because it represents a significant period in the emergence and development of large language models, retrieval-augmented generation, and learning from human feedback.

The search strategy employed combinations of keywords relevant to the scope of the study, including: *large language models, LLMs, reasoning, chain-of-thought, zero-shot reasoning, retrieval-augmented generation, RAG, knowledge grounding, instruction tuning, RLHF, human feedback, direct preference optimization, DPO, LLM-as-a-Judge,* and *LLM-based evaluation*. These terms were combined using the Boolean operators AND and OR, with the search syntax adapted to the requirements and functionality of each database.

3.2. Inclusion and Exclusion Criteria:

Studies were included if they met the following criteria: (1) they were published between 2020 and August 2026; (2) they had undergone peer review; (3) they directly addressed large language models; and (4) they were related to at least one of the four RRAJ dimensions, with sufficient information available regarding the study methodology, findings, and scientific contribution.

Conversely, non-peer-reviewed preprints for which no final published version was available were excluded, as were blog posts, commercial reports, journalistic materials, duplicate records, and studies in which LLMs were used only incidentally in an applied setting without making a substantive contribution relevant to the scope of this review. When multiple versions of the same study were available, the final published version was used.

3.3. Study Selection and Data Extraction:

The search and study selection process was guided by the general principles of the PRISMA 2020 guidelines, beginning with the identification of retrieved records, followed by duplicate removal, title and abstract screening, full-text assessment, and, finally, the inclusion of eligible studies in the final analysis (Page et al., 2021). The number of studies identified, screened, excluded, and ultimately included will be reported in a PRISMA flow diagram once the final screening process has been completed.

For each included study, key information was extracted, including the author(s) and year of publication, the technique or method employed, the model or models examined, the datasets or benchmarks used, the evaluation metrics, and the principal findings. Each study was also classified according to its primary RRAJ dimension, while cross-dimensional relationships were recorded when a study combined reasoning with retrieval, alignment, or evaluation.

3.4. Analytical Approach:

The study employed a Comparative Thematic Analysis to organize the literature across the four RRAJ dimensions and to compare developments within each dimension according to the nature of the capability, the source of knowledge or signals, the role of human involvement, the evaluation approach, and the principal limitations. A temporal analysis was also conducted to identify major shifts across the study period, without assuming that the four dimensions evolved in a linear chronological sequence.

In the final stage, the findings of the comparative analysis were used to develop the RRAJ framework as an integrative analytical model for explaining the relationships among Reasoning, Retrieval, Alignment, and Judging. Particular attention was given to the transition from the development of relatively distinct capabilities toward what this study conceptualizes as Functional Convergence among the capabilities of large language models.

4. Comparative Analytical Findings:

This section presents the findings derived from the comparative analysis of the included studies, with the aim of tracing the major transformations in the evolution of large language models across the four dimensions of the RRAJ framework: Reasoning–Retrieval–Alignment–Judging. The analysis extends beyond a chronological account of the relevant techniques by focusing on the nature of the computational mechanisms employed, the point of intervention within the system, the functional transformation represented by each technique, and the extent to which each study relates to one or more dimensions of RRAJ. This perspective enables the analysis to move beyond a descriptive review

of individual studies toward an interpretation of how LLMs have evolved from relatively distinct capabilities into increasingly integrated and functionally interconnected systems.

Table 1 summarizes the key studies that shaped this developmental trajectory and compares them across six unified analytical dimensions: the study, study type, technique or mechanism, point of computational intervention, primary functional transformation, and functional interconnection within the RRAJ framework. This comparative matrix serves as the foundation for the subsequent detailed analysis of reasoning, retrieval and knowledge grounding, alignment and preference optimization, and the transition from generation to LLM-as-a-Judge.

Table 1. Comparative Analytical Matrix of Key Studies on the Evolution of Large Language Models within the RRAJ Framework

Study & Study Type	**Technique & Mechanism**	**Point of Computational Intervention**	**Functional Transformation Represented**	**Functional Interconnection within RRAJ**
A. Reasoning Studies				
Brown et al. (2020). Foundational empirical study	In-Context Learning and Few-shot Learning	Context level	Transition from task-specific fine-tuning to task adaptation through contextual information	Primarily associated with Reasoning, providing a contextual foundation for the subsequent development of reasoning mechanisms
Wei et al. (2022), Chain-of-Thought. Empirical study	Chain-of-Thought Prompting	Context and inference-time reasoning	Transition from direct answer generation to the production of an intermediate reasoning trajectory	Primarily associated with Reasoning
Kojima et al. (2022). Empirical study	Zero-shot Chain-of-Thought	Context and inference-time reasoning	Elicitation of latent reasoning capabilities without requiring prior reasoning exemplars	Primarily associated with Reasoning

Zelikman et al. (2022), STaR. Empirical study	Bootstrapped Reasoning	Model parameters through iterative training	Transformation of model-generated reasoning trajectories into signals that are iteratively reused to improve model training	Demonstrates functional integration between Reasoning and Judging through the selection and refinement of reasoning trajectories before their reuse in training
Wang et al. (2023), Self-Consistency. Empirical study	Sampling with Consistency-Based Selection	Inference-time computation	Transition from reliance on a single reasoning trajectory to generating multiple trajectories and selecting the most consistent answer	Demonstrates functional integration between Reasoning and Judging, as consistency across reasoning trajectories is used to select the final output
Zhou et al. (2023), Least-to-Most. Empirical study	Problem Decomposition	Context and inference-time reasoning	Decomposition of a complex problem into an interconnected sequence of simpler subproblems	Primarily associated with Reasoning by structuring the problem-solving process into multiple stages
Yao et al. (2023), ReAct. Empirical study	Reasoning combined with Acting	Inference-time reasoning with an interface for interacting with tools or the environment	Linking the model's internal reasoning process to external actions and the dynamic acquisition of additional information	Demonstrates functional integration between Reasoning and Retrieval by using externally obtained information to update the reasoning trajectory
Gao et al. (2023), PAL. Empirical study	Program-Aided Language Models	Inference-time reasoning with an external	Separation of solution-plan generation from the execution of precise	Extends Reasoning through the use of an external execution tool, without constituting a

		program executor	computational operations	separate RRAJ dimension
Chen et al. (2023), Program of Thoughts. Empirical study	Program-of-Thoughts	Inference-time reasoning with program execution	Separation of semantic and linguistic reasoning from computational operations executed programmatically	Extends Reasoning by integrating it with external programmatic execution
Yao et al. (2023), Tree of Thoughts. Empirical study	Search over Thought States	Inference-time search	Transition from a linear reasoning chain to search across multiple alternatives, with the possibility of evaluation and backtracking	Demonstrates functional integration between Reasoning and Judging, as intermediate reasoning states are evaluated to determine subsequent search paths
Yang et al. (2025). Analytical empirical study	Long-CoT combined with Supervised Fine-Tuning, Reinforcement Learning, and Inference Scaling	Model parameters and inference-time computation	Transition from prompt-induced reasoning toward trained extended reasoning that incorporates revision, backtracking, and correction while scaling computational resources at time	Demonstrates functional integration among Reasoning, Alignment, and Judging through the combination of post-training, reward signals, and extended reasoning with iterative revision
B. Retrieval and Knowledge Grounding				
Guu et al. (2020), REALM.	Retrieval-Augmented Pre-training	Pre-training combined with external memory	Incorporation of retrieval from external knowledge	Primarily associated with Retrieval and Knowledge Grounding by linking model

Foundational empirical study			sources into the pre-training process	learning to an external knowledge memory
Karpukhin et al. (2020), DPR. Empirical study	Dense Passage Retrieval	External retrieval layer based on dense representations	Transition from lexical matching to semantic passage retrieval	Primarily associated with Retrieval and Knowledge Grounding, providing the retrieval infrastructure required for evidence-supported generative systems
Lewis et al. (2020), RAG. Foundational empirical study	Retrieval-Augmented Generation using a Retriever and Generator	External memory integrated into the generation process	Integration of knowledge encoded in model parameters with externally retrievable knowledge during generation	Primarily associated with Retrieval and Knowledge Grounding, with external evidence directly incorporated into the response-generation process
Izacard and Grave (2021), Fusion-in-Decoder. Empirical study	Fusion-in-Decoder	Retrieval combined with the decoding stage	Transition from reliance on a single passage to integrating information distributed across multiple retrieved passages	Demonstrates functional interconnection between Retrieval and evidence-based Reasoning by integrating multiple sources during answer generation
Borgeaud et al. (2022), RETRO. Empirical study	Retrieval-Enhanced Transformer	Model architecture combined with large-scale external memory	Use of external memory to expand the knowledge available to the system without a comparable increase in model parameter count	Primarily associated with Retrieval and Knowledge Grounding, supporting prediction through large-scale non-parametric memory

Izacard et al. (2023), Atlas. Empirical study	Few-shot Retrieval-Augmented Language Model	Training combined with updatable external memory	Decoupling updates to external knowledge from updates to model parameters while supporting learning from a limited number of examples	Demonstrates functional interconnection between Retrieval and contextual Reasoning by combining Few-shot Learning with access to external knowledge
Asai et al. (2024), Self-RAG. Empirical study	Adaptive Retrieval combined with Reflection Tokens	Model parameters, inference-time reasoning, and external retrieval	Transition from static retrieval to adaptive retrieval performed when needed, while evaluating the relevance of retrieved evidence and generated outputs	Demonstrates functional integration among Retrieval, Reasoning, and Judging by combining external knowledge retrieval with self-evaluation of evidence and outputs
Shi et al. (2024), REPLUG. Empirical study	Retrieval-Augmented Black-Box Language Models	Context combined with a trainable external retriever	Addition of retrieval capabilities to an existing language model without modifying its internal architecture or retraining the model itself	Primarily associated with Retrieval and Knowledge Grounding, supporting generation with external knowledge in a black-box setting
Li et al. (2026), SC-RAG. Empirical study	Self-Corrective Chain-of-Thought combined with Hybrid Retrieval	Retrieval and inference-time computation	Transition from merely providing evidence to assessing its relevance, resolving conflicts, and correcting its use	Demonstrates functional integration among Retrieval, Reasoning, and Judging through evidence-aware reasoning and self-correction mechanisms

			during answer generation	
C. Alignment and Preference Optimization				
Stiennon et al. (2020). Foundational empirical study	Reward Modeling combined with Reinforcement Learning from Human Feedback (RLHF)	Post-training stage	Transformation of human judgments about output quality into a learnable reward signal that can be used to optimize model behavior	Demonstrates functional integration between Alignment and Judging by using human judgments to train a reward model that guides subsequent model optimization
Wei et al. (2022), FLAN. Empirical study	Instruction Tuning	Model parameters during post-training	Transition from general text completion to instruction following and generalization to tasks not encountered during fine-tuning	Primarily associated with Alignment by training the model to respond more consistently to diverse instructions
Ouyang et al. (2022), InstructGPT. Foundational empirical study	Supervised Fine-Tuning combined with Reward Modeling and Proximal Policy Optimization	Post-training stage	Establishment of RLHF as an integrated pipeline for aligning model behavior with user intent and preferences	Demonstrates functional integration between Alignment and Judging, as human comparisons and reward modeling are used to steer model behavior toward preferred responses
Yuan et al. (2023), RRHF. Empirical study	Ranking Loss	Post-training stage	Simplification of the alignment process through direct	Demonstrates functional interconnection between Alignment and Judging

			learning from response rankings rather than relying entirely on the conventional reinforcement learning pipeline	by using response-quality rankings as a direct signal for updating the model
Rafailov et al. (2023), DPO. Empirical study	Direct Preference Optimization (DPO)	Post-training stage	Transition from the conventional pipeline based on a separate reward model and explicit reinforcement learning to direct policy optimization using preference data	Primarily associated with Alignment through the direct use of preference pairs to optimize model behavior
Chung et al. (2024). Large-scale empirical study	Scaled Instruction Fine-Tuning	Model parameters during post-training	Demonstration that scaling instruction fine-tuning and incorporating reasoning-related data can improve generalization to unseen tasks	Demonstrates functional interconnection between Alignment and Reasoning by incorporating reasoning-intensive data into the instruction-tuning process
Ethayarajh et al. (2024), KTO. Empirical study with a theoretical foundation	Human-Aware Loss using Kahneman–Tversky Optimization	Post-training stage	Transition from reliance on preferred–dispreferred response pairs to the use of simpler signals that	Primarily associated with Alignment by transforming simplified human signals into a direct optimization objective for model behavior

			distinguish desirable from undesirable outputs	
D. LLM-as-a-Judge				
Liu et al. (2023), G-Eval. Empirical study	Chain-of-Thought combined with Form-Filling Evaluation	Evaluation layer during inference	Transition from traditional reference-matching metrics to semantic evaluation based on a large language model and predefined evaluation criteria	Demonstrates functional interconnection between Judging and Reasoning by using structured reasoning steps to arrive at an evaluative judgment
Zheng et al. (2023), MT-Bench and Chatbot Arena. Benchmark-based empirical study	Pairwise LLM Judgment	Independent evaluation layer	Establishment of large language models as evaluators for comparing responses and conversational models	Demonstrates functional interconnection between Judging and Alignment, as evaluation relies substantially on human quality preferences and comparative criteria
Kim et al. (2024), Prometheus. Empirical study	Fine-tuned Rubric-based Judge	Independent specialized evaluator	Transition from reliance on a general-purpose closed judge model to an open, specialized evaluation model guided by explicit rubrics	Primarily associated with Judging through specialization of the model for fine-grained evaluation tasks based on clearly defined criteria

Wang et al. (2024), PandaLM. Empirical study	Fine-tuned Pairwise Judge	Independent specialized evaluator	Transformation of evaluation capability into a trainable skill implemented in a dedicated and relatively smaller model	Demonstrates functional interconnection between Judging and Alignment by learning evaluative judgments from assessment data aligned with human preferences
Wang et al. (2024), Large Language Models Are Not Fair Evaluators. Empirical study with bias analysis	Pairwise Judgment combined with Calibration	Evaluation layer	Shift from focusing solely on judgment accuracy toward examining evaluator fairness and reliability and mitigating biases such as position bias	Primarily associated with Judging, with particular emphasis on judgment reliability and calibration
Chiang et al. (2024), Chatbot Arena. Evaluation-platform-based empirical study	Blind Pairwise Human Preference	External human evaluation layer	Transition from static benchmark-based evaluation to dynamic assessment based on pairwise comparisons and actual user preferences	Demonstrates functional interconnection between Judging and Alignment, as human preferences serve as a reference criterion for assessing model quality
Zhu et al. (2025), JudgeLM. Empirical study	Fine-tuned Scalable Judge	Independent specialized evaluator	Development of scalable judge models of varying sizes while investigating and	Demonstrates functional interconnection between Judging and Alignment, as judge

			mitigating multiple sources of evaluation bias	training relies on judgment data designed to align model evaluations with reference assessments
Li et al. (2025). Analytical review	Taxonomy of Scoring, Ranking, and Selection	Meta-evaluation level	Conceptualization of LLM-as-a-Judge as a distinct evaluation paradigm encompassing multiple forms of judgment	Primarily associated with Judging, providing a theoretical synthesis of its mechanisms, evaluation modes, and relationships with broader model-evaluation practices
Gu et al. (2026). Systematic analytical review	Reliability-oriented LLM-as-a-Judge	Meta-evaluation level	Shift in research emphasis from demonstrating the feasibility of using LLMs as evaluators toward examining their reliability, consistency, and calibratability	Demonstrates functional interconnection among Judging, Alignment, and Reasoning by analyzing the relationships among judgment mechanisms, preference criteria, and the reasoning processes used to reach evaluative decisions.

The analytical table 1 presented a reveal that the evolution of large language models has not followed a single technical trajectory. Rather, it has involved a progressive expansion in the points of computational intervention within the overall system. In earlier work, improvements were concentrated primarily at the level of context and the model's internal capabilities, as illustrated by In-Context Learning and Chain-of-Thought prompting. Retrieval-oriented research subsequently introduced an external knowledge layer based on retrievers, indexes, and non-parametric memory. Alignment techniques, by contrast, shifted part of the optimization process to the post-training stage through instructions, preferences, and reward signals, while LLM-as-a-Judge research added a further,

relatively independent layer for evaluating model outputs. This progression suggests that the primary unit of development is no longer the model in isolation, but rather the integrated computational architecture surrounding and interacting with it.

The matrix also indicates a parallel transformation in the sources of knowledge and control signals used by LLM-based systems. Models have moved from relying predominantly on knowledge encoded in their parameters toward incorporating contextual information, retrieved evidence, human preferences, evaluation criteria, and judgments generated by other models. This shift represents more than the addition of new information sources; it reflects a redistribution of functional responsibilities across the system. Retrieval determines which knowledge enters the generation process, alignment influences which behaviors are preferred, and the judge provides signals regarding whether an output should be accepted, ranked, or revised. Consequently, overall system quality increasingly depends on the quality of the signals exchanged among system components, rather than solely on the capabilities of the underlying language model.

The comparison further highlights an increasing tendency toward functional convergence across the RRAJ dimensions. Whereas earlier studies often focused on a single capability, more recent approaches increasingly combine multiple functions. ReAct, for example, connects reasoning with interaction with external information sources; Self-RAG and SC-RAG integrate retrieval, reasoning, and evaluation; and RLHF and RRHF link preference evaluation to subsequent alignment procedures. Such interconnections suggest that the boundaries among Reasoning, Retrieval, Alignment, and Judging are becoming progressively less distinct. The technical trajectory is therefore shifting from the development of isolated components toward systems in which the output of one stage can guide, trigger, or reinitiate another stage.

Finally, the matrix demonstrates that increasing integration is accompanied by a corresponding shift in the nature of the associated challenges. While early research primarily emphasized performance improvement, reliability, computational cost, and error propagation across components have become increasingly central concerns. Multi-path reasoning increases inference-time computational demands; retrieval introduces risks related to the relevance, quality, and reliability of evidence; alignment depends on the quality and representativeness of preference signals; and judge models may introduce their own biases into the broader system loop. The comparison therefore suggests that future progress may depend less on simply adding further components and more on managing the interactions among them, diagnosing sources of error, and determining when each component should be activated and at what computational cost. This finding provides the empirical and conceptual basis for the subsequent development of the RRAJ computational-analytical framework.

5. Computational-Analytical Framework:

The preceding analysis demonstrates that Reasoning, Retrieval, Alignment, and Judging differ not only in their functional roles, but also in the computational point at which intervention occurs, the type of signal employed, the nature of the memory involved, the optimization mechanism, and the stage at which each technique operates, whether during training or inference. Accordingly, this study extends RRAJ from a functional taxonomy into a multilayer computational-analytical framework that enables LLM techniques to be compared in terms of system architecture and execution mechanisms, rather than function alone.

5.1. Computational Architecture of the RRAJ Framework:

At an abstract level, the operation of an RRAJ-based system can be represented as a set of interconnected components. The process begins with an input (x) and a context (C), after which the reasoning component produces a representation or reasoning state (z). If the task requires external knowledge, (z), or part of it, is used to formulate a retrieval query (q), which is then used to retrieve a set of evidence (D) from an external memory (M). The model subsequently uses these elements to generate a response (y) under a set of alignment constraints (A). Finally, the judging component (J) assigns a judgment or score (s) to the response according to a specified criterion or rubric.

The resulting computational cycle can be expressed in simplified form as follows:

Input → Reasoning State → Retrieval Query → Evidence → Aligned Generation → Judgment

followed by:

Judgment → Re-reason / Re-retrieve / Regenerate

In this formulation, Judging does not necessarily represent the terminal stage of the process. Rather, the resulting judgment can function as a control signal that redirects, reactivates, or modifies earlier components of the system, thereby enabling iterative reasoning, retrieval, and generation.

5.2. Technical Layers Mechanisms across the Four RRAJ Dimensions:

The four dimensions of RRAJ do not operate through identical computational mechanisms. Reasoning is typically implemented within the decoding and inference process, whereas Retrieval relies primarily on components external to the core language model, such as retrievers, indexes, and external memory. Alignment is predominantly implemented during the post-training stage through mechanisms that modify model behavior using instructions, preferences, or reward signals. Judging, in contrast, may be performed either by a separate evaluator model or by the same model through self-evaluation mechanisms. These differences highlight that the four RRAJ dimensions occupy distinct, yet increasingly interconnected, computational layers within contemporary LLM-based systems.

Table 2. Computational Layers and Execution Mechanisms within the RRAJ Framework

RRAJ Dimension	Key Computational Techniques	Point of Computational Intervention	Type of Memory or Signal Used	Primary Stage
Reasoning	Chain-of-Thought, Zero-shot Chain-of-Thought, Sampling, Self-Consistency, Search, Tree of Thoughts, Tool Use, Inference-time Scaling	Decoding stage and reasoning control module	Context, together with intermediate states and reasoning traces	Inference-time reasoning
Retrieval and Knowledge Grounding	Sparse Retrieval, Dense Retrieval, Dual Encoders, Embeddings, Vector Indexing, Approximate Nearest Neighbor Search, Hybrid Retrieval, Re-ranking, Evidence Fusion, Adaptive Retrieval	External knowledge-access interface	External non-parametric memory	Training and inference time
Alignment and Preference Optimization	Supervised Fine-Tuning, Instruction Tuning, Reward Modeling, Proximal Policy Optimization, RRHF, Direct Preference Optimization, Kahneman–Tversky Optimization	Model parameters during post-training	Demonstrations, human preferences, and reward signals	Post-training stage
Judging	Pointwise Scoring, Pairwise Comparison, Ranking, Rubric-based Evaluation, Reference-based Evaluation, Reference-free Evaluation, Calibration, Multi-Judge Aggregation	Evaluator or critic layer	Evaluation criteria, preferences, and reference answers	Inference-time evaluation

The table demonstrates that RRAJ does not represent four computationally homogeneous techniques. Reasoning and Judging rely heavily on inference-time computation, whereas Alignment is more closely associated with modifying model parameters during post-training. Retrieval, by contrast, introduces an external non-parametric memory into the system.

This distinction is analytically important because it suggests that the evolution of LLMs has occurred through an expansion of the locations at which computation is performed: from model parameters, to contextual processing, then to external memory, followed by additional inference-time computation, and ultimately to evaluation and feedback layers. From this perspective, advances in LLMs can be understood not only as improvements in individual techniques, but also as a progressive redistribution of computation across different components and stages of the overall system.

5.3. Levels of Computational Intervention and Integration within the RRAJ Framework

The analysis reveals that RRAJ techniques do not operate at a single location within a large language model system; rather, they are distributed across four principal computational levels. The first is Parameter-level Intervention, as exemplified by Instruction Tuning, Supervised Fine-Tuning, RLHF, and DPO, where model weights are modified during post-training to alter model behavior. The second is Context-level Intervention, as in Few-shot Prompting and Chain-of-Thought, where model behavior is influenced through instructions, demonstrations, and contextual information without modifying the model parameters. The third level is External-memory Intervention, implemented through techniques such as RAG, dense retrieval, vector indexes, and evidence fusion, enabling the model to access external knowledge that is not directly encoded in its parameters. The fourth level involves Inference-time Intervention, as illustrated by Self-Consistency, Tree of Thoughts, Long-CoT, retry mechanisms, and self-evaluation, where additional computational resources are allocated during task execution to improve the resulting output.

These levels do not necessarily operate independently; rather, they exchange information and control signals across the components of RRAJ. Reasoning may reveal a knowledge gap that triggers retrieval, while retrieved evidence may subsequently redirect the reasoning process. Alignment techniques influence how the model formulates and constrains its responses, whereas the judging layer may use retrieved evidence or preference criteria to assess output quality. When the evaluator identifies a deficiency, the response can be routed back to an earlier stage for renewed reasoning, retrieval, or generation. In this sense, evaluation becomes a feedback mechanism rather than merely a terminal stage.

The comparison across studies suggests three patterns of computational integration. The first consists of Isolated Capabilities, in which each function operates largely independently. The second comprises

Coupled Systems, which combine two or more functions, such as reasoning with retrieval or retrieval with evaluation. The most advanced pattern is represented by Adaptive Closed-loop Systems, which can determine whether retrieval, further reasoning, evaluation, or correction is required according to the state and demands of the task, rather than executing the same sequence of stages in every case.

Accordingly, system design shifts from a fixed pipeline in which a prompt is received and a response is generated toward a more adaptive architecture that can be summarized as follows: task planning, reasoning, retrieval when needed, guided generation, evaluation, and correction when necessary. Within the RRAJ framework, this transition is conceptualized as Adaptive Orchestration, whereby efficiency does not depend on activating all components at all times, but rather on determining which component should be invoked, when it should be activated, and how much computational effort the task requires.

This perspective enables techniques to be compared along a limited set of shared computational dimensions, most notably: the point of intervention within the system, execution stage, requirement for model-parameter modification, use of external memory, nature of the feedback signal, capacity for reevaluation and correction, and computational cost. In this way, the analysis moves beyond comparing individual algorithmic labels toward examining how knowledge, computation, and control are distributed throughout the system.

On this basis, the principal technical implication of the RRAJ framework is that the evolution of large language models does not merely reflect a progression across reasoning, retrieval, alignment, and judging. Rather, it represents a transition from a monolithic language model relying primarily on parametric knowledge and direct generation toward a multi-component computational system in which knowledge, computation, control mechanisms, and evaluation are distributed across interconnected components. The analytical value of RRAJ therefore lies not only in explaining what large language models are capable of doing, but also in identifying where these capabilities are implemented within the system, how their components interact, and how evaluation and feedback can be used to adaptively reorganize system behavior.

6. Discussion:

The analytical findings across the four dimensions indicate that the evolution of large language models during the study period cannot be explained through a single trajectory based solely on increasing model scale or improving generative capacity. The more significant transformation lies in the shift from LLMs that rely primarily on knowledge encoded in model parameters and direct generation toward composite computational systems in which reasoning, knowledge access, behavioral control, and evaluation are distributed across multiple components and stages.

From this perspective, the RRAJ framework enables the literature to be interpreted as reflecting a broader transition from Scaling to Functional and Computational Integration.

6.1. From Scaling the Model to Scaling the System

The early stages of LLM development focused extensively on the relationship between increases in model size, training data, and the capabilities that emerged as a result. However, the findings of this study suggest that subsequent improvements have increasingly depended on factors beyond parameter scaling alone. Some approaches allocate additional computational resources during inference, others introduce external memory through Retrieval, while Alignment techniques rely on human-derived or preference-based signals to modify model behavior. LLM-as-a-Judge approaches, in turn, introduce an evaluation layer that can operate without necessarily modifying the underlying generator model.

This finding is consistent with the broader direction of the literature, which suggests that the evolution of LLM architectures increasingly involves a growing set of training, adaptation, orchestration, and evaluation components rather than merely scaling the core model itself (Shao et al., 2024). Accordingly, the concept of capability in LLM-based systems can be reformulated from:

Model Capability

to:

System Capability = Model + Context + Memory + Inference + Feedback + Evaluation

This reformulation implies that comparing two LLM-based systems solely on the basis of parameter count is becoming increasingly insufficient for explaining meaningful differences in their actual capabilities and performance.

6.2. Distribution of Knowledge, Computation, and Control:

The comparison across the RRAJ dimensions reveals three parallel transformations: the distribution of knowledge, the distribution of computation, and the distribution of control.

With respect to knowledge, information is no longer confined to parametric memory but is distributed across model weights, contextual information, and externally retrieved sources. In terms of computation, computational resources are no longer concentrated primarily in the training stage; instead, there is increasing reliance on inference-time computation through the generation of multiple reasoning trajectories, search, revision, and retry mechanisms. Behavioral control, likewise, is no longer determined solely by pretraining data, but is increasingly shaped by instructions, preferences, reward signals, and evaluation criteria.

This distribution provides one of the key explanations for the increased flexibility of contemporary LLM-based systems. At the same time, however, it expands the number of potential failure points

within the overall system. Errors may originate from incorrect reasoning, irrelevant or misleading retrieval, biased preference signals, or unreliable judge models. Consequently, system reliability increasingly depends not only on the robustness of individual components, but also on the quality of their interactions and the mechanisms used to detect and contain errors across components.

6.3. Functional Convergence:

The literature indicates that Reasoning, Retrieval, Alignment, and Judging can no longer be regarded as entirely independent research trajectories. Modern retrieval systems increasingly incorporate mechanisms for assessing evidence quality and initiating additional retrieval when necessary. Evaluation systems, meanwhile, may employ reasoning mechanisms to produce more interpretable and structured judgments, while judgments generated by LLM-based evaluators can themselves be transformed into preference data for subsequent Alignment.

Accordingly, this study proposes Functional Convergence as a more accurate characterization of contemporary LLM evolution than a strictly linear chronological progression. The most advanced stage of development does not necessarily involve the emergence of a new capability that replaces an earlier one; rather, it is characterized by an increasing density of functional connections among existing capabilities.

For example:

Reasoning + Retrieval enables evidence-grounded reasoning.

Retrieval + Judging enables outputs to be evaluated against retrieved evidence.

Judging + Alignment enables evaluative judgments to be transformed into preference or training signals.

Reasoning + Judging enables revision, critique, and iterative retry mechanisms.

When these interactions operate recursively, the system moves beyond a linear pipeline toward a closed-loop architecture, in which the output of a later stage can reactivate, redirect, or modify an earlier stage. Such architectures therefore represent a shift from sequential task execution toward dynamically coordinated systems in which reasoning, retrieval, alignment, and evaluation mutually influence one another.

6.4. The Changing Role of Humans in the LLM Development Cycle:

The findings also reveal an important transformation in the position of humans within the LLM ecosystem. In earlier stages, humans primarily provided data or examples; they subsequently assumed the role of instruction providers, and later supplied comparisons and preferences used for Alignment.

In LLM-as-a-Judge systems, however, the human role increasingly shifts toward designing evaluation rubrics, calibrating evaluators, and verifying their reliability.

This progression can be summarized as follows:

Example Provider → Instructor → Preference Provider → Evaluator → Judge Calibrator / Supervisor

This transition does not imply the disappearance of human involvement, but rather a shift toward a higher level of supervisory control. As evaluation becomes increasingly automated, the human role becomes more closely associated with defining evaluation criteria, auditing bias, calibrating automated evaluators, and handling cases in which machine-generated judgments cannot be considered sufficiently reliable.

This issue becomes particularly important as LLM-as-a-Judge approaches continue to expand. Recent literature indicates that consistency, bias, contextual sensitivity, and judgment-design choices remain fundamental challenges, making human supervision and evaluator calibration essential for the development of reliable automated evaluation systems (Gu et al., 2026).

6.5. Judging as a Turning Point in the LLM Lifecycle:

Judging represents more than simply the fourth dimension of RRAJ because it introduces into the system the capability to assess the quality of its own outputs or those produced by other systems. The LLM-as-a-Judge paradigm has expanded to encompass scoring, ranking, selection, and evaluation according to diverse criteria, reflecting a broader transition in which LLMs move from being objects of evaluation to becoming active participants in the evaluation process itself (Li et al., 2025).

From the perspective of this study, the significance of Judging lies in its potential to generate a signal that can be propagated back to earlier stages of the system. If the evaluator determines that a response is insufficiently supported by evidence, Retrieval can be reactivated; if a logical error is detected, the system can initiate another Reasoning cycle; and if the problem concerns behavior, style, or preference conformity, Generation can be repeated under different Alignment constraints.

The process therefore shifts from: **Generate → Evaluate**

to: **Generate → Evaluate → Diagnose → Correct**

This transition provides a conceptual basis for understanding the emerging trajectory toward self-evaluating and self-correcting LLMs as a consequence of convergence among the four RRAJ dimensions rather than as an independent line of development.

Taken together, the findings support an interpretation of LLM evolution as a transition from stand-alone generative models to composite cognitive-computational systems. In such systems, knowledge is distributed across multiple sources, computation is distributed across different stages of operation,

behavioral control is shaped by multiple forms of signals, and evaluation itself becomes an integral component of the operational cycle.

Accordingly, the principal contribution of the RRAJ framework does not lie in arranging four techniques into a sequential taxonomy. Rather, it lies in demonstrating that the future development of LLMs is increasingly determined by the coordination of relationships among reasoning, knowledge access, preference alignment, and evaluation.

6.6. Illustrative Application Case: The RRAJ Cycle in Answering an Educational Question:

To illustrate how the proposed framework operates in practice, consider a user who asks an LLM-based system the following question:

"Why do we see lightning before we hear thunder? Explain it simply and cite a reliable scientific source."

Although the question appears straightforward, it imposes several requirements that the system must address simultaneously: identifying the causal relationship between lightning and thunder, verifying the relevant scientific information, presenting the explanation in accessible language, and ensuring that the final response is both accurate and consistent with the user's request. This example demonstrates how the RRAJ dimensions can move from a theoretical classification to an interconnected computational workflow.

First, the Reasoning stage.

The system semantically analyzes the request and decomposes it into three sub-tasks: explaining the phenomenon, simplifying the explanation, and providing a reliable source. The reasoning process identifies that lightning and thunder originate from essentially the same atmospheric event, but light and sound propagate at substantially different speeds. The relevant variable to explain is therefore the difference in arrival time rather than a difference in the time at which the two phenomena occur. Computationally, this stage takes place within the context and inference-time reasoning process, where the model transforms a general natural-language question into a causal structure that can guide response construction.

Second, the Retrieval and Knowledge Grounding stage.

Because the user explicitly requests a reliable scientific source, relying solely on the model's parametric memory is not the most appropriate strategy. The system therefore converts the informational need identified during the previous stage into a retrieval query focused on the relationship between the speed of light, the speed of sound, and the delayed perception of thunder. The retriever searches a reliable scientific or educational source, after which the retrieved passages are

ranked according to their relevance using semantic retrieval and re-ranking mechanisms. The objective is not to retrieve the greatest possible amount of information, but to identify evidence that supports the specific causal claim: light reaches the observer far more rapidly than sound, and therefore lightning is seen before thunder is heard.

This example highlights the direct relationship between Reasoning and Retrieval. Retrieval is not initiated arbitrarily; rather, it is triggered after Reasoning has identified the knowledge gap that requires verification. The retrieved information is then returned to the model context, where it constrains the generation process to claims that are supported by the available evidence.

Third, the Alignment stage.

Scientific correctness alone is insufficient because the user has requested a "simple" explanation.

Alignment therefore governs the form, style, and linguistic level of the response. Rather than providing an extended physical explanation involving electromagnetic waves and acoustic propagation, the system might formulate the response as follows: *Lightning and thunder occur at almost the same time, but light travels much faster than sound. As a result, the light from the lightning reaches your eyes first, while the sound of thunder reaches your ears later.* A concise scientific source can then be provided.

Alignment thus serves a function that is distinct from Retrieval. Retrieval addresses the question "What information is correct?", whereas Alignment addresses "How should this information be presented to this user?" At the computational level, the system draws on the instructions contained in the prompt as well as behavioral patterns acquired through Instruction Tuning and Preference Optimization to balance scientific accuracy with accessibility.

Fourth, the Judging stage.

Before the response is accepted, an evaluation layer can assess it using a question-specific rubric comprising four criteria: scientific accuracy, consistency between the claim and the retrieved source, adherence to the requested level of simplicity, and completeness with respect to all elements of the user's request. If the response is scientifically correct but excessively technical, the deficiency can be classified as an Alignment problem. If it is simple but contains a claim that is unsupported by the retrieved evidence, the problem can be classified as a Knowledge Grounding failure. If the response confuses the physical cause of thunder with the reason thunder is heard later, the error can instead be attributed to Reasoning.

The importance of this stage lies in the fact that the Judge need not merely assign a general score such as "good" or "poor." Rather, the judgment can be used to diagnose the location of failure within

the RRAJ cycle. The nature of the identified error can then determine the next computational action: a knowledge-related error triggers renewed retrieval, a causal error triggers another reasoning cycle, failure to satisfy the requested level of simplicity triggers regeneration under revised Alignment constraints, whereas a response that satisfies all criteria can terminate the cycle and be delivered to the user.

The application flow can therefore be represented conceptually and technically as follows:

User Question → Requirement Analysis → Causal Explanation Construction → Identification of Verification Need → Scientific Evidence Retrieval → Evidence Re-ranking → Simplified Response Generation → Evaluation of Accuracy, Grounding, and Instruction Adherence → Response Acceptance or Reactivation of the Component Responsible for the Error

This case illustrates that the principal value of the RRAJ framework does not lie in routing every question through four fixed stages. Rather, it lies in the dynamic orchestration of different functions according to the requirements of the task. Reasoning determines what needs to be established, Retrieval provides the necessary evidence, Alignment transforms that knowledge into a response appropriate for the user, and Judging functions as a monitoring and diagnostic layer capable of redirecting the system toward the component responsible for a detected deficiency. RRAJ therefore represents a transition from a model centered primarily on Prompt-to-Response Generation toward a system capable of understanding the task, accessing relevant knowledge, adapting the response, evaluating its quality, and correcting it when necessary.

7. Challenges and Open Issues:

Despite the progress made in integrating the components of the RRAJ framework, several important challenges remain unresolved. These include:

- Error propagation across Reasoning, Retrieval, Alignment, and Judging, which may make it difficult to identify the true source of a system failure.
- The generation of reasoning trajectories that appear linguistically coherent and logically plausible without necessarily representing a correct or reliable causal process.
- Dependence on external evidence that may be outdated, inaccurate, incomplete, or inconsistent with the model's internal knowledge.
- The variability of human preferences across users, contexts, cultures, and domains, which substantially increases the complexity of Alignment.
- The susceptibility of LLM-as-a-Judge systems to judgment biases, prompt formulation, answer ordering, and response length, potentially reintroducing evaluation errors into the broader system cycle.

- The risk of error amplification in systems where the same model is used for generation, evaluation, and correction without an independent verification mechanism.
- Increased computational cost and response latency resulting from the combined use of extended reasoning, retrieval, evaluation, and repeated correction cycles.
- The difficulty of applying RRAJ in the medical domain without rigorous evidence verification, uncertainty estimation, and expert human oversight before any recommendation is accepted or acted upon.
- The heightened level of risk associated with military and security applications because of information sensitivity and the potential for misinformation or evaluative bias, thereby requiring clearly defined constraints on system autonomy.
- The need in legal and financial domains for continuous verification of source currency and the relevant regulatory context before system outputs can be considered reliable.
- The difficulty of assigning responsibility when failures occur in multi-component systems, making traceability and accountability essential system requirements.
- The continuing challenge of defining the appropriate boundaries of autonomy and determining when a decision should be escalated to human oversight, particularly in high-stakes applications.

8. Future Research Agenda:

Based on the findings of this study and the challenges identified above, several directions emerge as priorities for future research:

- Developing more verifiable reasoning mechanisms, such that quality is assessed not only by the correctness of the final answer but also by the extent to which the evidence and intermediate steps supporting the decision can be independently verified.
- Developing adaptive retrieval mechanisms capable of determining when external knowledge is required, identifying which sources are most reliable, and appropriately handling conflicting, incomplete, or insufficient evidence.
- Developing Alignment approaches that accommodate variation in preferences, contexts, and values rather than assuming a single uniform preference structure across all users and applications.
- Improving the reliability of LLM-as-a-Judge through calibration, uncertainty estimation, multi-judge aggregation, and human verification for high-risk decisions.
- Moving from evaluation toward diagnosis and self-correction, such that the system does not merely determine that an output is incorrect, but also identifies the source of the failure and whether renewed reasoning, retrieval, or generation is required.

- Developing adaptive orchestration across RRAJ components so that only the necessary component is activated according to task characteristics, confidence level, risk, and computational cost.
- Investigating the addition of an Execution/Action Layer after evaluation in applications that do not terminate with a textual response, including robotics, autonomous agents, and control systems, so that an accepted judgment can be translated into an executable action within the environment.
- Developing pre-execution verification mechanisms that prevent reasoning or evaluation errors from being translated directly into physical actions, particularly in robotics, medicine, industrial systems, and military applications.
- Investigating post-execution feedback loops in which the observed outcome of an action is compared with the expected state, triggering renewed reasoning, evaluation, or correction when deviations are detected.
- Defining the boundaries between human and machine autonomy, including the degree of authority that may appropriately be delegated to a system according to application risk and the reversibility of its decisions.

Accordingly, the future trajectory of the framework may extend beyond Reasoning, Retrieval, Alignment, and Judging toward a broader cycle that incorporates evaluation, execution, monitoring of action outcomes, and subsequent correction. Such an extension becomes particularly important as large language models evolve from systems that primarily generate knowledge and decisions into systems capable of exerting direct effects on the physical environment.

9. Conclusion:

This study aimed to provide an integrative interpretation of the evolution of large language models during the period 2020–2026, based on the observation that recent literature has moved beyond a primary focus on generative capability toward a broader set of interconnected cognitive, behavioral, and evaluative functions. Through a systematic analysis of the literature, the study examined four principal dimensions: Reasoning, Retrieval and Knowledge Grounding, Alignment and Preference Optimization, and LLM-as-a-Judge.

The analysis demonstrated that the evolution of LLMs cannot be adequately explained as a linear chronological progression or as a direct consequence of increasing model scale alone. Rather, the field has gradually shifted from reliance on parametric knowledge and direct generation toward systems in which knowledge, computation, and control are distributed across model parameters, contextual information, external memory, inference-time reasoning mechanisms, preference signals, and evaluation processes. Accordingly, the most appropriate unit of analysis is no longer the model in isolation, but the integrated computational system within which the model operates.

Within this context, the study introduced the RRAJ framework as an analytical perspective that connects the four dimensions and examines them at two complementary levels. At the functional level, the framework clarifies the role performed by each capability. At the computational level, it identifies the point of intervention, the source of knowledge or signals, the nature of the memory involved, the stage of execution, and the mechanisms through which feedback is incorporated. The analysis indicates that one of the clearest trends in the literature is the transition from relatively isolated capabilities toward Functional Convergence, whereby the boundaries among Reasoning, Retrieval, Alignment, and Judging are becoming increasingly interconnected within contemporary LLM-based systems.

The findings further suggest that incorporating evaluation into the operational cycle of LLMs represents an important transition. In principle, evaluation enables judgment to move beyond its role as a terminal stage and become a signal capable of triggering renewed reasoning, retrieval, or generation. This creates a pathway from LLM-as-a-Judge toward increasingly self-evaluating and self-correcting LLMs. However, this trajectory is accompanied by substantial challenges, including error propagation across components, the limited reliability of some reasoning trajectories, the quality and relevance of retrieved evidence, variability in human preferences, biases in judge models, and increasing computational cost.

On this basis, the study argues that the most important future trajectory does not lie in improving each component independently, but in developing systems capable of adaptively coordinating Reasoning, Retrieval, Alignment, and Judging. This requires a transition from fixed processing pipelines toward systems that can determine when retrieval is necessary, when reasoning should be expanded, when external evaluation is required, and when a task should be retried or escalated to human oversight. From this perspective, the future development of large language models is likely to depend increasingly on the effectiveness with which knowledge, computation, control, and evaluation are orchestrated across the system as a whole.

10. Declaration:

The author affirms that the research is original and authored by the author. The author also confirms that there is no funding source for the study and no conflict of interest. The author confirms the use of artificial intelligence tools for proofreading and linguistic editing.

11. References

Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. (2024). Self-RAG: Learning to retrieve, generate, and critique through self-reflection. In *International Conference on Learning Representations (ICLR)*.

Borgeaud, S., Mensch, A., Hoffmann, J., Cai, T., Rutherford, E., Millican, K., Van Den Driessche, G. B., Lespiau, J.-B., Damoc, B., Clark, A., De Las Casas, D., Guy, A., Menick, J., Ring, R., Hennigan, T., Huang, S., Maggiore, L., Jones, C., Cassirer, A., … Sifre, L. (2022). Improving language models by retrieving from trillions of tokens. In *Proceedings of the 39th International Conference on Machine Learning* (Vol. 162, pp. 2206–2240). PMLR.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

Chen, W., Ma, X., Wang, X., & Cohen, W. W. (2023). Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks. *Transactions on Machine Learning Research*.

Chiang, W.-L., Zheng, L., Sheng, Y., Angelopoulos, A. N., Li, T., Li, D., Zhu, B., Zhang, H., Jordan, M., Gonzalez, J. E., & Stoica, I. (2024). Chatbot Arena: An open platform for evaluating LLMs by human preference. In *Proceedings of the 41st International Conference on Machine Learning* (Vol. 235, pp. 8359–8388). PMLR.

Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., Brahma, S., Webson, A., Gu, S. S., Dai, Z., Suzgun, M., Chen, X., Chowdhery, A., Castro-Ros, A., Pellat, M., Robinson, K., … Wei, J. (2024). Scaling instruction-finetuned language models. *Journal of Machine Learning Research, 25*(70), 1–53.

Ethayarajh, K., Xu, W., Muennighoff, N., Jurafsky, D., & Kiela, D. (2024). Model alignment as prospect theoretic optimization. In *Proceedings of the 41st International Conference on Machine Learning* (Vol. 235, pp. 12634–12651). PMLR.

Gao, L., Madaan, A., Zhou, S., Alon, U., Liu, P., Yang, Y., Callan, J., & Neubig, G. (2023). PAL: Program-aided language models. In *Proceedings of the 40th International Conference on Machine Learning* (Vol. 202, pp. 10764–10799). PMLR.

Gu, J., Jiang, X., Shi, Z., Tian, H., Zhai, X., Xu, C., Li, W., Shen, Y., Ma, S., Liu, H., Wang, S., Zhang, K., Lin, Z., Zhang, B., Ni, L. M.-S., Gao, W., Wang, Y., & Guo, J. (2026). A survey on LLM-as-a-judge. *The Innovation, 7*(6), 101253. https://doi.org/10.1016/j.xinn.2025.101253

Guu, K., Lee, K., Tung, Z., Pasupat, P., & Chang, M. (2020). Retrieval augmented language model pre-training. In *Proceedings of the 37th International Conference on Machine Learning* (Vol. 119, pp. 3929–3938). PMLR.

Izacard, G., & Grave, E. (2021). Leveraging passage retrieval with generative models for open domain question answering. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. 874–880). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.eacl-main.74

Izacard, G., Lewis, P., Lomeli, M., Hosseini, L., Petroni, F., Schick, T., Dwivedi-Yu, J., Joulin, A., Riedel, S., & Grave, E. (2023). Atlas: Few-shot learning with retrieval augmented language models. *Journal of Machine Learning Research, 24*(251), 1–43.

Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W.-t. (2020). Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 6769–6781). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.550

Kim, M. J., Pertsch, K., Karamcheti, S., Xiao, T., Balakrishna, A., Nair, S., Rafailov, R., Foster, E. P., Sanketi, P. R., Vuong, Q., Kollar, T., Burchfiel, B., Tedrake, R., Sadigh, D., Levine, S., Liang, P., & Finn, C. (2025). OpenVLA: An open-source vision-language-action model. In *Proceedings of the 8th Conference on Robot Learning* (Vol. 270, pp. 2679–2713). PMLR.

Kim, S., Shin, J., Cho, Y., Jang, J., Longpre, S., Lee, H., Yun, S., Shin, S., Kim, S., Thorne, J., & Seo, M. (2024). Prometheus: Inducing fine-grained evaluation capability in language models. In *International Conference on Learning Representations (ICLR)*.

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. *Advances in Neural Information Processing Systems, 35*, 22199–22213. https://doi.org/10.52202/068431-1613

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459–9474.

Li, D., Jiang, B., Huang, L., Beigi, A., Zhao, C., Tan, Z., Bhattacharjee, A., Jiang, Y., Chen, C., Wu, T., Shu, K., Cheng, L., & Liu, H. (2025). From generation to judgment: Opportunities and challenges of LLM-as-a-judge. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing* (pp. 2757–2791). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.emnlp-main.138

Li, Y., Ke, W., Liu, J., Wang, P., Liu, J., & He, Y. (2026). Towards evidence-aware retrieval-augmented generation via self-corrective chain-of-thought. *Information Processing & Management, 63*(2), 104369. https://doi.org/10.1016/j.ipm.2025.104369

Liu, Y., Iter, D., Xu, Y., Wang, S., Xu, R., & Zhu, C. (2023). G-Eval: NLG evaluation using GPT-4 with better human alignment. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 2511–2522). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.153

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35*, 27730–27744. https://doi.org/10.52202/068431-2011

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., & Finn, C. (2023). Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems, 36*, 53728–53741. https://doi.org/10.52202/075280-2338

Shao, M., Basit, A., Karri, R., & Shafique, M. (2024). Survey of different large language model architectures: Trends, benchmarks, and challenges. *IEEE Access, 12*, 188664–188706. https://doi.org/10.1109/ACCESS.2024.3482107

Shi, W., Min, S., Yasunaga, M., Seo, M., James, R., Lewis, M., Zettlemoyer, L., & Yih, W.-t. (2024). REPLUG: Retrieval-augmented black-box language models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (pp. 8371–8384). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.463

Stiennon, N., Ouyang, L., Wu, J., Ziegler, D. M., Lowe, R., Voss, C., Radford, A., Amodei, D., & Christiano, P. F. (2020). Learning to summarize with human feedback. *Advances in Neural Information Processing Systems, 33*, 3008–3021.

Wang, P., Li, L., Chen, L., Cai, Z., Zhu, D., Lin, B., Cao, Y., Kong, L., Liu, Q., Liu, T., & Sui, Z. (2024). Large language models are not fair evaluators. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 9440–9450). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.511

Wang, X., Wei, J., Schuurmans, D., Le, Q. V., Chi, E. H., Narang, S., Chowdhery, A., & Zhou, D. (2023). Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations (ICLR)*.

Wang, Y., Yu, Z., Yao, W., Zeng, Z., Yang, L., Wang, C., Chen, H., Jiang, C., Xie, R., Wang, J., Xie, X., Ye, W., Zhang, S., & Zhang, Y. (2024). PandaLM: An automatic evaluation benchmark for LLM instruction tuning optimization. In *International Conference on Learning Representations (ICLR)*.

Wei, J., Bosma, M., Zhao, V. Y., Guu, K., Yu, A. W., Lester, B., Du, N., Dai, A. M., & Le, Q. V. (2022). Finetuned language models are zero-shot learners. In *International Conference on Learning Representations (ICLR)*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E. H., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems, 35*, 24824–24837. https://doi.org/10.52202/068431-1800

Yang, S., Tong, Y., Niu, X., Neubig, G., & Yue, X. (2025). Demystifying long chain-of-thought reasoning. In *Proceedings of the 42nd International Conference on Machine Learning* (Vol. 267, pp. 71177–71209). PMLR.

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems, 36*, 11809–11822. https://doi.org/10.52202/075280-0517

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR)*.

Yuan, H., Yuan, Z., Tan, C., Wang, W., Huang, S., & Huang, F. (2023). RRHF: Rank responses to align language models with human feedback. *Advances in Neural Information Processing Systems, 36*, 10935–10950. https://doi.org/10.52202/075280-0482

Zelikman, E., Wu, Y., Mu, J., & Goodman, N. D. (2022). STaR: Bootstrapping reasoning with reasoning. *Advances in Neural Information Processing Systems, 35*, 15476–15488.

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *Advances in Neural Information Processing Systems, 36*, 46595–46623. https://doi.org/10.52202/075280-2020

Zhou, D., Schärli, N., Hou, L., Wei, J., Scales, N., Wang, X., Schuurmans, D., Cui, C., Bousquet, O., Le, Q. V., & Chi, E. H. (2023). Least-to-most prompting enables complex reasoning in large language models. In *International Conference on Learning Representations (ICLR)*.

Zhu, L., Wang, X., & Wang, X. (2025). JudgeLM: Fine-tuned large language models are scalable judges. In *International Conference on Learning Representations (ICLR)*.

Zhuge, M., Zhao, C., Ashley, D. R., Wang, W., Khizbullin, D., Xiong, Y., Liu, Z., Chang, E., Krishnamoorthi, R., Tian, Y., Shi, Y., Chandra, V., & Schmidhuber, J. (2025). Agent-as-a-Judge: Evaluate agents with agents. In *Proceedings of the 42nd International Conference on Machine Learning* (Vol. 267, pp. 80569–80611). PMLR.

Doi: https://doi.org/10.52132/Ajrsp.e.2026.88.1