Reasoning Systems as Structured Processes: Foundations, Failures, and Formal Criteria

Saleh Nikooroo and Thomas Engel
saleh.nikooroo@uni.lu, thomas.engel@uni.lu

Abstract—This paper outlines a general formal framework for reasoning systems, intended to support future analysis of inference architectures across domains. We model reasoning systems as structured tuples comprising phenomena, explanation space, inference and generation maps, and a principle base. The formulation accommodates logical, algorithmic, and learning-based reasoning processes within a unified structural schema, while remaining agnostic to any specific reasoning algorithm or logic system. We survey basic internal criteria—including coherence, soundness, and completeness—and catalog typical failure modes such as contradiction, incompleteness, and non-convergence. The framework also admits dynamic behaviors like iterative refinement and principle evolution. The goal of this work is to establish a foundational structure for representing and comparing reasoning systems, particularly in contexts where internal failure, adaptation, or fragmentation may arise. No specific solution architecture is proposed; instead, we aim to support future theoretical and practical investigations into reasoning under structural constraint.

Index Terms—Reasoning systems, formal structure, inference dynamics, failure modes, coherence, soundness, adaptive reasoning.

I. Introduction

Reasoning systems are typically formalized within well-established frameworks such as symbolic logic, optimization theory, or machine learning architectures. These paradigms often operate under the assumptions of full internal consistency, total deductive closure, or the global applicability of inference rules. While these assumptions enable tractable and mathematically elegant systems, they can obscure or fail to accommodate the real-world characteristics of reasoning: partial information, structural fragmentation, or evolving principle sets.

This paper departs from traditional models by proposing a general framework for reasoning systems as structured, modular entities. Crucially, our formulation does not presuppose any particular reasoning algorithm or deductive paradigm; it aims instead to describe the structural conditions under which reasoning may emerge or fail.

Rather than focusing solely on logical deduction or optimization performance, we examine reasoning systems as dynamic processes that interpret phenomena, generate explanations, and validate them internally against a set of governing principles.

Our motivation stems from the observation that many reasoning failures—from logical paradoxes to deadlocks in constrained systems—are not anomalies but structural symptoms of rigid or insufficient internal frameworks. To better understand these limits and potentials, we seek a formalism that:

- Captures the internal composition of a reasoning system, including its inputs, outputs, generative maps, and principle base;
- Enables a classification of failure modes—such as contradiction, incompleteness, or non-convergence—within a unified framework;
- Provides structural criteria for internal coherence, inference validity, and explanation sufficiency;
- Views reasoning as an iterative and adaptive structural process, rather than a static rule-based engine.

The remainder of this paper proceeds as follows. In Section II, we define a minimal formal structure for reasoning systems. Section III introduces key notions of coherence, soundness, and completeness. Section IV categorizes failure modes and their structural implications. Section V examines the internal dynamics of reasoning, including fixed-point behavior and principle drift. Section VI instantiates the formalism in several reasoning domains. We conclude with a brief discussion of implications and future directions.

II. Formal Structure of a Reasoning System

We define a reasoning system as a structured quintuple:

$$\mathcal{R} = (P, E, f, g, \Pi)$$

where each component plays a distinct functional role:

- P is the set of **phenomena**, inputs, or observed problems that the system is intended to interpret or solve;
- E is the **explanation space**, consisting of candidate solutions, hypotheses, or structured outputs;
- $f : P \rightarrow E$ is the **inference map**, producing explanations from phenomena;
- $g : E \rightarrow P$ is the **generation map**, reconstructing or predicting phenomena from explanations;
- Π is the **principle system**, a set of structural, logical, or epistemic constraints that govern both f and g.

This formulation is intentionally agnostic to implementation. For instance:

- In symbolic logic, P may be a set of premises, E a set of theorems, f a deduction operator, and Π a proof system;
- In constrained optimization, P is the space of objective functions and constraints, E the solution space, f a solver, and Π the feasibility conditions;
- In a neural inference setting, f may be a learned function approximator, while Π encodes architectural or regularization constraints.

We emphasize that f and g are not necessarily inverses, nor are they guaranteed to be bijective or even total. In

fact, the limits of their definability and mutual consistency will be a central concern throughout this work.

The role of Π is to constrain the admissibility and structural behavior of both f and g. It may encode axioms, domain-specific rules, inductive biases, or dynamic constraints that evolve over time. Importantly, a reasoning system may be internally well-defined (i.e., f, g exist and operate) while still violating Π, or failing to apply Π consistently.

In what follows, we use this model to analyze the internal structure, coherence, and failure modes of reasoning systems in a general and implementation-independent way.

III. COHERENCE, SOUNDNESS, AND VALIDITY

Having established the formal structure of a reasoning system $\mathcal{R} = (P, E, f, g, \Pi)$, we now define the core internal criteria by which such a system may be evaluated. These include:

- **Coherence**: Whether the explanations produced by f and their reconstructions via g are consistent with the original inputs.
- **Soundness**: Whether all generated explanations respect the governing principles Π.
- **Completeness**: Whether the system is capable of producing valid explanations for all admissible phenomena in P.

A. Coherence

We define a reasoning system as *coherent* if, for all $p \in P$ such that $f(p)$ is defined,

$$g(f(p)) \approx p$$

under a chosen notion of approximation (e.g., exact equality, semantic closeness, or acceptable reconstruction error). Coherence captures the ability of a reasoning system to reconstruct or validate its own interpretive steps.

Perfect coherence implies $g \circ f = \mathrm{id}_P$ on the domain where f is defined. In practice, this is often relaxed to a tolerable discrepancy, especially in probabilistic or learned systems.

B. Soundness

A reasoning system is *sound* if all explanations produced by f are consistent with the principles in Π. That is,

$$\forall p \in P, \quad f(p) \models \Pi$$

where $\models$ denotes logical or structural satisfaction. Soundness guarantees that inference respects internal rules, whether logical axioms, physical constraints, or algorithmic feasibility.

Soundness may also be localized: some systems are sound only with respect to subsets of Π, or under specific operational contexts.

C. Completeness

We say the system is *complete* if, for every admissible $p \in P$, there exists a valid explanation $e \in E$ such that:

$$f(p) = e \quad \text{and} \quad e \models \Pi$$

This ensures that the reasoning system can handle its full intended domain without structural blind spots.

Note that completeness is not merely the existence of $f(p)$ but also its adherence to Π. A system may be able to compute $f(p)$ for all p, yet remain incomplete if those explanations violate Π.

D. Fixed-Point Interpretations

In certain systems, a fixed-point relationship may serve as an idealization:

$$f(g(e)) = e \quad \text{and/or} \quad g(f(p)) = p$$

Such relationships reflect ideal coherence under perfect invertibility. In general, systems approximate rather than achieve these fixed points.

E. Joint Evaluation

The interplay between coherence, soundness, and completeness reveals much about the system's structural health:

- A system can be sound but incoherent (e.g., it reasons correctly but fails to reconstruct inputs).
- It can be coherent but incomplete (e.g., reconstructs reliably but fails on many inputs).
- Achieving all three properties simultaneously is rare and often requires tightly constrained Π or system simplification.

These criteria provide a framework for analyzing where reasoning systems succeed or break down, setting the stage for the failure typologies to follow.

IV. TYPOLOGY OF FAILURE IN REASONING SYSTEMS

Despite a system's formal structure and internal constraints, reasoning processes often encounter breakdowns. These failures are not necessarily flaws in implementation but rather indicators of structural insufficiency, misalignment, or rigidity within the system. We now outline a typology of such failures, categorized by their causes and manifestations.

A. Contradiction

A contradiction occurs when the output of the reasoning process violates the governing principles Π. That is, for some $p \in P$:

$$f(p) \not\models \Pi$$

This can arise from incompatible axioms, overloaded inference rules, or misaligned generative behavior in g. Contradictions indicate internal inconsistency and may lead to epistemic collapse or rejection of the explanation.

B. Incompleteness

A system is said to exhibit incompleteness when it fails to provide explanations for certain phenomena within its intended scope. Formally, there exists $p \in P$ such that $f(p)$ is undefined or inadmissible:

$$\exists p \in P \quad \text{such that} \quad f(p) \notin E \quad \text{or} \quad f(p) \models \neg \Pi$$

Incompleteness can stem from under-specified principles, rigid constraint boundaries, or unanticipated problem instances.

C. Non-Convergence

Some reasoning systems involve iterative or recursive processes. A failure mode emerges when these iterations do not converge to a stable explanation. For example, repeated application of $f \circ g$ or $g \circ f$ might yield divergent or oscillating results:

$$\lim_{n \to \infty} (f \circ g)^n(e) \quad \text{does not exist}$$

Non-convergence is common in optimization-based, heuristic, or neural reasoning frameworks.

D. Overfitting and Underfitting

When f over-specializes to training or observed instances, the system may become brittle and fail to generalize—this is overfitting. Conversely, if f is too coarse or regularized, it may produce vague or non-informative explanations—this is underfitting.

These behaviors often arise from misaligned Π, overly flexible or overly constrained function classes, or improper selection of inductive biases.

E. Structural Deadlock

A reasoning system may become structurally inert when its internal logic is self-consistent but incapable of progressing in the face of novel or ambiguous inputs. This deadlock occurs when:

- f is defined but constant or non-responsive across large regions of P;
- Π prohibits admissible alternatives due to over-constraint;
- or f outputs trivial explanations (e.g., $e = \texttt{null}$ or tautologies).

Such deadlocks are particularly insidious, as the system appears functional but fails to engage meaningfully with its problem space.

F. Failure Summary

These failure types are not mutually exclusive. A reasoning system may simultaneously suffer from contradiction in some regions, incompleteness in others, and deadlock elsewhere. The typology above provides a diagnostic vocabulary for characterizing and comparing reasoning systems in practice.

In the following section, we explore how such systems evolve internally—sometimes recovering from failure, sometimes reinforcing it.

V. INTERNAL DYNAMICS AND REASONING EVOLUTION

Reasoning systems are not always static entities. In many applications, they operate over time—through iterative refinement, self-correction, or evolving structural constraints. This section explores the internal dynamics of such systems, focusing on how they respond to error, adapt to novelty, or restructure themselves without external intervention.

A. Sequential Inference and Iterative Structure

In systems where f or g are defined recursively or in stages, reasoning proceeds through a sequence of internal steps:

$$e_0 = f(p), \quad e_1 = f(g(e_0)), \quad \ldots, \quad e_n = f(g(e_{n-1}))$$

This chain may converge, cycle, or diverge depending on the nature of f, g, and Π. The goal of such iterations may include refining an explanation, validating internal coherence, or escaping suboptimal initial mappings.

B. Error-Driven Adjustment

Some systems adjust their internal mappings in response to discrepancies between predicted and observed phenomena:

$$\delta_p = p - g(f(p))$$

Such error signals can trigger refinement of f, tuning of g, or revision of Π. This dynamic is especially prominent in adaptive learning systems, where gradient-based or rule-based updates aim to reduce reconstruction error or improve inference fidelity.

C. Principle Drift

Over time, the principle set Π may itself evolve:

$$\Pi_0 \to \Pi_1 \to \cdots \to \Pi_t$$

This evolution may be triggered by contradictions, poor performance, or the emergence of new problem domains. Principle drift alters the admissibility conditions and effectively changes the structure of the reasoning system. It reflects a meta-level response: rather than repairing f or g, the system alters what it considers valid.

D. Self-Regularization

Some reasoning systems include built-in mechanisms to avoid or correct undesirable behavior. These may include:

- Penalizing incoherent mappings (e.g., regularization terms in optimization);
- Constraining search space to prevent overfitting;
- Disabling unstable regions in f or g through gating or pruning mechanisms.

These structural safeguards help enforce stability and steer the system toward valid reasoning behavior without explicit supervision.

E. Local vs. Global Adaptation

Adaptation may occur locally (e.g., only for specific phenomena $p \in P$) or globally (altering f, g, or Π across the entire system). Systems that support local updates may be more resilient but risk fragmentation; global adaptations offer coherence at the risk of rigidity.

F. Failure Response Modes

Not all systems respond to failure. Some ignore error entirely; others collapse or halt; others adapt. The presence or absence of failure-response dynamics—especially principle drift and error correction—can be used to classify systems into static vs. evolving reasoning architectures.

In the next section, we instantiate this framework with examples drawn from logic, optimization, and learning systems.

VI. EXAMPLES OF REASONING SYSTEMS

We now illustrate the general framework by instantiating it in three distinct domains: deductive logic, constrained optimization, and structured neural inference. These examples demonstrate the flexibility of the (P, E, f, g, Π) formulation and how different reasoning paradigms fit within its structure.

A. Example 1: Deductive Logic System

- **Phenomena** (P): Sets of premises or assumptions.
- **Explanation space** (E): Theorems or derived propositions.
- **Inference map** (f): A derivation function applying inference rules (e.g., modus ponens).
- **Generation map** (g): Reconstructs minimal premises or antecedents from a given theorem (where applicable).
- **Principles** (Π): Axioms, inference rules, and proof constraints (e.g., propositional logic axioms).

Coherence in this setting corresponds to whether derived conclusions can be traced back to accepted premises. *Soundness* ensures that all derivations respect logical axioms. *Completeness* refers to whether all logically entailed theorems can be reached.

B. Example 2: Constrained Optimization Solver

- **Phenomena** (P): Problem specifications—objective functions and constraint sets.
- **Explanation space** (E): Candidate solutions or configurations.
- **Inference map** (f): Optimization routine mapping problems to solutions.
- **Generation map** (g): Reconstructs problem structure from candidate solutions (e.g., via duality).
- **Principles** (Π): Feasibility conditions, KKT constraints, or convexity assumptions.

Here, *coherence* implies that optimal solutions explain the posed problem faithfully. *Soundness* requires that outputs satisfy all constraints. *Completeness* reflects the solver's ability to find feasible solutions across the problem domain.

C. Example 3: Structured Neural Inference

- **Phenomena** (P): Input data points or observations (e.g., images, text).
- **Explanation space** (E): Feature embeddings, latent codes, or predicted labels.
- **Inference map** (f): Neural network performing encoding or classification.

- **Generation map** (g): Decoder or generative model reconstructing input.
- **Principles** (Π): Inductive biases encoded by architecture, loss functions, or regularization terms.

Coherence relates to how well reconstructions match inputs. *Soundness* reflects whether predictions respect the model's inductive assumptions. *Completeness* captures whether the model generalizes across the full data distribution.

D. Cross-Example Summary

Each of the above systems instantiates the same general structure but emphasizes different components of reasoning:

- Logic emphasizes Π (axioms) and f (deduction).
- Optimization emphasizes f and g under feasibility constraints.
- Neural inference emphasizes learned f, approximate g, and implicit Π via architecture.

This diversity underscores the versatility of the formal framework, which abstracts away from domain specifics to analyze structural and functional integrity at a higher level.

In the concluding section, we summarize our contributions and suggest directions for further research.

VII. RELATED WORK

Structured Reasoning Frameworks and Epistemic Models

Recent work moves beyond classical logic by modeling reasoning as structured, modular, or coherence-driven rather than truth-functional. Simon [1] develops a coherence-based framework for biased reasoning, emphasizing internal consistency as the main organizing principle—resonant with our model's allowance for structural failure and belief reconfiguration.

Casini [2] provides a flexible logic for conditional reasoning beyond truth-functional semantics, supporting the kind of partial inference and non-monotonicity that our framework tolerates explicitly. Arieli [3] classifies argumentation schemes via postulate satisfaction, proposing inferential formalisms that account for contradiction without collapse, conceptually aligning with our treatment of epistemic tensions.

Gärdenfors [4] offers a geometric model of conceptual reasoning, and Kido [5] presents probabilistic symbolic frameworks for structured inference—both motivating our belief graph's topological encoding of epistemic structure. Paulino-Passos [6] interprets explanation as a non-monotonic reasoning process, suggesting evolving inference paths; by contrast, we permit persistent contradiction to co-exist with structural coherence.

Bushuev [7] discusses multimodal reasoning via TRIZ principles, prioritizing adaptability and contradiction management—reflecting our interest in epistemic resilience rather than deductive purity.

Modular, Compositional, and Iterative Architectures

The modular design of reasoning architectures has gained traction. Christianos et al. [8] propose Pangu-Agent, a generalist agent embedding reasoning modules for improved

generalization. Creswell [9] builds modular chains of fine-tuned neural units to enforce reasoning faithfulness and reduce hallucination, paralleling our interest in structured yet flexible inference models. SCREWS [10] introduces a reasoning framework that supports revision and modular recombination of inference steps. WanJun et al. [11] disentangle representation and reasoning through a compositional transformer model, while Zhou et al. [12] introduce SELF-DISCOVER to enable LLMs to autonomously compose reasoning chains—both reinforcing structural modularity over end-to-end black-box reasoning. Fu [13] offers MORSE, a dynamic modular reasoning framework tailored to explanation generation. Hua et al. [14] combine neural and symbolic reasoning pathways in a two-system model, advocating for cognitive complementarity—a principle echoed in our separation of operational and constraint mappings.

Topological and Graph-Based Representations

Graphical and topological reasoning models provide an interpretive scaffold. Zhang et al. [15] propose Diagram of Thought (DoT), using Topos Theory to model LLM reasoning as a directed acyclic graph. Ho et al. [16] frame logical traversal as dialectical dialogue trees tolerant to inconsistency. Zhu et al. [17] propose structural representation learning to generalize across reasoning types.

Our own belief graphs continue this lineage but shift emphasis from entailment to structural coherence—allowing persistent contradiction and cluster-based epistemics.

Meta-Reasoning, Verification, and Dynamic Strategy Control

Sui [18] proposes Meta-Reasoner, a dynamic controller for strategy selection during inference time. Xiang [19] extends this to meta chain-of-thought supervision, enhancing internal traceability. Raza [20] explores logical task verification using solvers, while ReCEval [21] evaluates reasoning chains for correctness and informativeness. Both highlight meta-level oversight and structural inspection—similar in spirit to our separation of confidence versus credibility within belief networks. Ling et al. [22] offers a deductive framework supporting self-verification and chain decomposition, reinforcing reasoning transparency. Wei [23] proposes AlignRAG, a feedback-based framework for aligning multi-hop retrieval chains—conceptually related to our epistemic feedback model.

Inductive, Causal, and Failure-Oriented Models

Qiu et al. [24] analyze inductive hypothesis refinement, revealing limitations in rule application and generalization—an issue our model reframes as structural incoherence. Wang et al. [25] examine transformer generalization failures as indicators of latent reasoning limits. Tang et al. [26] present CausalGPT, a multi-agent architecture for causality-aware reasoning. Saparov [27] offers a formal critique of LLM proof planning, demonstrating accurate micro-inference but poor global coherence. Jung et al. [28] improve reasoning robustness through recursive, logically consistent explanations, reinforcing our emphasis on iterative reasoning integrity. Prystawski [29] proposes that

reasoning arises from local correlations among observed variables—a structural insight compatible with our model's cluster-driven reasoning across graph components.

General Syntheses and Reasoning Surveys

Sun et al. [30] survey reasoning with foundation models, covering techniques across multi-modal, multi-agent, and structured reasoning. Plaat [31] consolidates prompting and stepwise reasoning in LLMs, identifying current gaps in systematic reasoning. Saied [32] surveys decision-making frameworks, emphasizing modular decision support—underscoring the need for structural substrates like belief graphs. Wang et al. [33] and Radhakrishnan et al. [34] develop Chain-of-Knowledge prompting and decomposition-based reasoning, respectively, both contributing to the understanding of structured deduction.

Summary

Across this landscape, a convergence emerges: toward frameworks that prioritize internal structure, revision, and epistemic transparency. Whether through modular reasoning units, topological encoding, or verification protocols, the field is moving away from opaque statistical inference toward explainable, compositional, and resilient systems. Our proposed belief graph framework fits within this shift, offering a contradiction-tolerant and structurally explicit substrate that separates reasoning integration from source reliability and procedural logic.

VIII. CONCLUSION AND OUTLOOK

This paper has proposed a general framework for reasoning systems, formalized as structured entities composed of phenomena, explanations, inference and generation mappings, and principle sets. We have shown that this formulation:

- Captures diverse reasoning paradigms—including logic, optimization, and learning—under a unified structural model;
- Defines internal evaluation criteria such as coherence, soundness, and completeness;
- Supports a detailed typology of failure modes, including contradiction, incompleteness, non-convergence, and deadlock;
- Enables the study of internal dynamics, from iterative refinement and error correction to evolving principles and self-regularization.

By decoupling reasoning from any specific implementation or representational formalism, the model invites cross-domain analysis of structural reasoning properties. It offers a foundation for diagnosing reasoning pathologies, comparing systems, and guiding principled design.

Future Directions

This work opens several avenues for future exploration:

- **Modular Composition**: How can reasoning systems be composed from subsystems while preserving or enhancing structural integrity?
- **Resilience and Repair**: Can local adaptations (e.g., error-driven refinement) ensure global soundness and coherence?

- **Dynamic Principle Systems**: What governs the evolution of Π in adaptive or self-revising systems?
- **Evaluation Metrics**: Beyond coherence and soundness, what metrics capture the robustness, generality, or strategic capacity of a reasoning system?

While we have not addressed interaction between distinct reasoning systems in this work, such considerations lie beyond the present scope and are left for future investigation.

In closing, the reasoning system framework offers a structural and dynamic view of inference architectures. It emphasizes not only what systems conclude, but how they operate, evolve, and fail. We believe this perspective will prove valuable across fields that rely on principled but adaptable reasoning processes.

REFERENCES

[1] D. Simon and S. J. Read, "Toward a general framework of biased reasoning: Coherence-based reasoning," *Perspectives on Psychological Science*, 2023.

[2] G. Casini and U. Straccia, "A general framework for modelling conditional reasoning - preliminary report," in *International Conference on Principles of Knowledge Representation and Reasoning*, 2022.

[3] O. Arieli *et al.*, "A postulate-driven study of logical argumentation," *Artificial Intelligence*, vol. 315, p. 103912, 2023.

[4] P. Gärdenfors and M. Osta-Vélez, "Reasoning with concepts: A unifying framework," *Minds and Machines*, 2023.

[5] H. Kido, "Generative logic models for data-based symbolic reasoning," in *International Workshop on Artificial Intelligence and Cognition*, 2022.

[6] G. Paulino-Passos and F. Toni, "On interactive explanations as non-monotonic reasoning," *arXiv preprint arXiv:2208.00316*, 2022.

[7] S. Bushuev *et al.*, "Reasoning mechanism in multimodal ai models based on the triz principles," *International Scientific Journal of Computing*, 2025. Forthcoming.

[8] F. Christianos *et al.*, "Pangu-agent: A fine-tunable generalist agent with structured reasoning," *arXiv preprint arXiv:2312.14878*, 2023.

[9] A. Creswell and M. Shanahan, "Faithful reasoning using large language models," *arXiv preprint arXiv:2208.14271*, 2022.

[10] K. Sridhar *et al.*, "Screws: A modular framework for reasoning with revisions," *arXiv preprint arXiv:2309.13075*, 2023.

[11] W. Zhong, "Disentangling reasoning capabilities from language models with compositional reasoning transformers," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2022.

[12] P. Zhou *et al.*, "Self-discover: Large language models self-compose reasoning structures," *Neural Information Processing Systems (NeurIPS)*, 2024.

[13] X. Fu, "Dynamic modularized reasoning for compositional structured explanation generation," *arXiv preprint arXiv:2307.08744*, 2023.

[14] W. Hua and Y. Zhang, "System 1 + system 2 = better world: Neural-symbolic chain of logic reasoning," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2022.

[15] Y. Zhang, "On the diagram of thought," *arXiv preprint arXiv:2403.00001*, 2024.

[16] L. T. T. Ho, "Dialogue-based explanations for logical reasoning using structured argumentation," *arXiv preprint arXiv:2502.06483*, 2025.

[17] Z. Zhu, "Learning representations for reasoning: Generalizing across diverse structures," *arXiv preprint arXiv:2402.07615*, 2024.

[18] Y. Sui, "Meta-reasoner: Dynamic guidance for optimized inference-time reasoning in large language models," *arXiv preprint arXiv:2503.10101*, 2025.

[19] V. Xiang, "Towards system 2 reasoning in llms: Learning how to think with meta chain-of-thought." arXiv preprint arXiv:2502.00001, 2025.

[20] M. Raza, "Instantiation-based formalization of logical reasoning tasks using language models and logical solvers," *arXiv preprint arXiv:2501.11234*, 2025.

[21] A. Prasad, "Receval: Evaluating reasoning chains via correctness and informativeness," in *Conference on Empirical Methods in Natural Language Processing*, 2023.

[22] Z. Ling *et al.*, "Deductive verification of chain-of-thought reasoning," *Neural Information Processing Systems*, 2023.

[23] J. Wei, "Alignrag: An adaptable framework for resolving misalignments in retrieval-aware reasoning of rag." arXiv preprint arXiv:2501.00001, 2025.

[24] L. Qiu *et al.*, "Phenomenal yet puzzling: Testing inductive reasoning capabilities of language models with hypothesis refinement," in *International Conference on Learning Representations (ICLR)*, 2023.

[25] B. Wang *et al.*, "Grokkinged transformers are implicit reasoners: A mechanistic journey to the edge of generalization," *arXiv preprint arXiv:2405.15071*, 2024.

[26] Z. Tang *et al.*, "Towards causalgpt: A multi-agent approach for faithful knowledge reasoning via promoting causal consistency in llms," *arXiv preprint arXiv:2308.11914*, 2023.

[27] A. Saparov and H. He, "Language models are greedy reasoners: A systematic formal analysis of chain-of-thought," in *International Conference on Learning Representations*, 2023.

[28] J. Jung *et al.*, "Maieutic prompting: Logically consistent reasoning with recursive explanations," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2022.

[29] B. Prystawski, "Why think step-by-step? reasoning emerges from the locality of experience," in *Neural Information Processing Systems*, 2023.

[30] J. Sun *et al.*, "A survey of reasoning with foundation models: Concepts, methodologies, and outlook," *ACM Computing Surveys*, 2023.

[31] A. Plaat, "Reasoning with large language models, a survey." arXiv preprint arXiv:2403.00001, 2024.

[32] W. M. Saied, "Comprehensive synthesis of decision-making in complex systems," in *International Conference on Computing and Information*, 2024.

[33] J. Wang *et al.*, "Boosting language models reasoning with chain-of-knowledge prompting," in *Annual Meeting of the Association for Computational Linguistics*, 2023.

[34] A. Radhakrishnan *et al.*, "Question decomposition improves the faithfulness of model-generated reasoning," *arXiv preprint arXiv:2307.11768*, 2023.