GRACE: Graph Reasoning with Adaptive Correlated Equilibrium

Khuong Nguyen

University of Virginia

Charlottesville, VA, USA

ABSTRACT

Large Language Models (LLMs) are prone to hallucination, a problem that standard Retrieval-Augmented Generation (RAG) mitigates but cannot eliminate due to its static, linear execution. Existing pipelines typically treat all queries identically, failing to adapt verification effort or abstain when retrieved evidence is noisy or irrelevant. To address this rigidity, we introduce **GRACE**, a framework that reconceptualizes RAG as a cooperative multi-agent game. GRACE coordinates the Retriever, Generator, and Verifier using a Correlated Equilibrium (CE) policy, allowing a central Mediator to recommend optimal joint strategies—such as deep verification or strategic abstention—conditioned on graph-theoretic trust signals. Experimental results on the HotpotQA benchmark demonstrate that GRACE functions as a "Safety Filter", effectively utilizing risk-aware coordination to screen out hallucinations and achieve significantly higher effective reliability than static baselines. However, the system also exhibits a too conservative behavior profile, governed by the variance of its structural input signals. Our findings validate game-theoretic control as a promising paradigm for robust, adaptive reasoning, managing the trade-off between recall and safety through dynamic multi-agent orchestration.

1 INTRODUCTION

Large Language Models (LLMs) have rapidly advanced the capabilities of natural language systems, enabling fluent and context-aware generation across a wide range of tasks. Despite their impressive performance, a persistent and fundamental challenge remains unresolved: LLMs frequently *hallucinate*, producing confident yet factually incorrect statements. These failures undermine reliability in applications that require grounded reasoning, factual precision, or verifiable decision-making.

A growing body of work suggests that hallucinations emerge not from accidental model flaws, but from structural pressures within the modern LLM training pipeline. As summarized in Figure 1, contemporary LLMs are developed via a three-stage process—*pre-training*, *fine-tuning*, and *alignment*—as described in the survey by Zhao et al. [20]. Each stage optimizes for linguistic fluency, instruction-following, or reward alignment, but none directly optimize for calibrated truthfulness. Consequently, LLMs are statistically pressured to guess under uncertainty, producing plausible but incorrect completions, a phenomenon rigorously analyzed by Kalai et al. [6].

To mitigate these issues, Retrieval-Augmented Generation (RAG) introduces external grounding by retrieving evidence from text corpora or knowledge graphs. While effective in principle, standard RAG pipelines follow a rigid sequence: retrieve k documents, append them to the prompt, and compel the model to generate an answer [9]. This static approach introduces several limitations:

(1) **Sensitivity to noisy or irrelevant evidence.** Retrieval errors can mislead the generator, contributing to failures such as the "lost-in-the-middle" phenomenon [11].

(2) **Inefficient computation.** Deep retrieval and robust verification (e.g., self-consistency) are computationally expensive but are typically applied uniformly across all queries.

(3) **Lack of abstention mechanisms.** Most systems are optimized to always answer, rather than to be reliably correct, and therefore lack principled ways to say "I do not know."

These limitations highlight the need for adaptive decision-making pipelines that condition their behavior on the quality and reliability of available evidence.

In this work, we introduce **GRACE** (Graph Reasoning with Adaptive Correlated Equilibrium), a framework that reconceptualizes retrieval-augmented reasoning as a cooperative multi-agent game. GRACE models the Retriever, Generator, and Verifier as strategic agents whose behaviors must be coordinated to balance accuracy, compute cost, and the risk of hallucination. We frame this interaction using a *correlated equilibrium* (CE) policy [1], which allows a central Mediator to recommend joint actions conditioned on shared signals. CE is well-suited for this setting because it enables correlated strategies-such as deep retrieval plus verification, or

shallow retrieval plus abstention-while remaining computationally tractable. The Mediator extracts graph-theoretic reliability signals, including PageRank-based node trustworthiness, evidence-path coherence, and subgraph diversity, and maps them to an optimal joint action such as deep retrieval, verification, or safe abstention. The full implementation of GRACE is available as open-source at:
`https://github.com/knguyen2000/grace`

2 BACKGROUND AND MOTIVATION

Although external retrieval mitigates some forms of hallucination, it does not address the deeper structural issue that retrieval-augmented systems must make a sequence of interdependent decisions under uncertainty. At each query, the system must implicitly determine how much evidence to gather, how strongly to trust parametric knowledge, whether the generated answer is sufficiently supported, and whether abstention is preferable to risking an incorrect response. Traditional RAG pipelines collapse these choices into a single, fixed pattern of execution, leaving no mechanism to adapt strategy based on the quality or reliability of the available evidence.

This rigidity becomes problematic in settings where retrieval quality varies widely across queries. Some questions admit shallow, high-confidence evidence; others require multi-hop reasoning over sparse or weakly connected graphs; still others cannot be reliably answered at all. Treating all queries identically leads either to excessive computation or to increased hallucination risk. This motivates viewing retrieval-augmented reasoning not merely as a retrieval problem, but as a structured decision-making task in which different actions incur different levels of cost and uncertainty. A natural perspective, therefore, is to treat the Retriever, Generator, and Verifier as components with distinct roles and failure modes that must be coordinated rather than executed in isolation. Their interactions resemble those of cooperative agents operating under a shared objective: maximize correctness while minimizing computational cost and the likelihood of unsupported answers. Game-theoretic frameworks, particularly those allowing a central coordinating mechanism, offer mathematical tools for modeling such joint decision processes.

CE, introduced by Aumann [1], provides exactly this capability. Unlike Nash equilibria, which constrain agents to act independently, CE allows a mediator to issue recommendations based on shared information, producing joint strategies that are both coherent and utility-maximizing. This property is essential in retrieval-augmented reasoning, where the usefulness of one component's action depends on the actions of the others: deep retrieval is only worthwhile if verification will follow; abstention is only rational when both retrieval and generation signal low confidence; verification is only valuable when the generator's output is unstable or weakly grounded. This perspective motivates GRACE, which leverages structural information from a knowledge graph to guide coordination among components. By extracting graph-theoretic signals such as node reliability, path coherence, and evidence diversity, the system can estimate the level of uncertainty surrounding each query and select a correlated joint action accordingly. This enables a form of adaptive computation in which expensive strategies are reserved for ambiguous cases, while simple queries are handled efficiently. In doing so, GRACE moves beyond static RAG pipelines and toward a principled framework for reliable, evidence-aware reasoning in LLM-based systems.

3 GRACE METHOD

GRACE reframes retrieval-augmented reasoning as a coordinated multi-agent decision problem, in which retrieval, generation, and verification are treated not as fixed pipeline stages but as strategic components whose behaviors must be jointly optimized. Traditional RAG systems implicitly assume that every query should follow the same retrieval depth, the same generation strategy, and the same verification procedure. In contrast, GRACE adapts its computation to the quality of available evidence, allocating effort only where uncertainty is high and abstaining where evidence is insufficient. This requires a mechanism that can (1) interpret structural signals from the knowledge graph, (2) coordinate multiple reasoning modules under a shared objective, and (3) produce joint actions that balance accuracy, computational cost, and hallucination risk.

To formalize this adaptive reasoning process, we introduce a multi-agent framework in which each component of the system is modeled as an agent with its own action space. Their interactions are regulated by a CE policy that selects a joint action conditioned on graph-theoretic reliability signals. We begin by describing the components of the GRACE framework.

3.1 Framework

GRACE consists of three functional agents—Retriever, Generator, and Verifier—and a central Mediator that coordinates their actions. Each agent is responsible for a distinct stage of the reasoning process, and each has access to a limited set of actions with different computational costs and reliability implications. This design allows the system to choose between fast but weaker strategies and slower but more robust ones depending on the structural cues extracted from the knowledge graph.

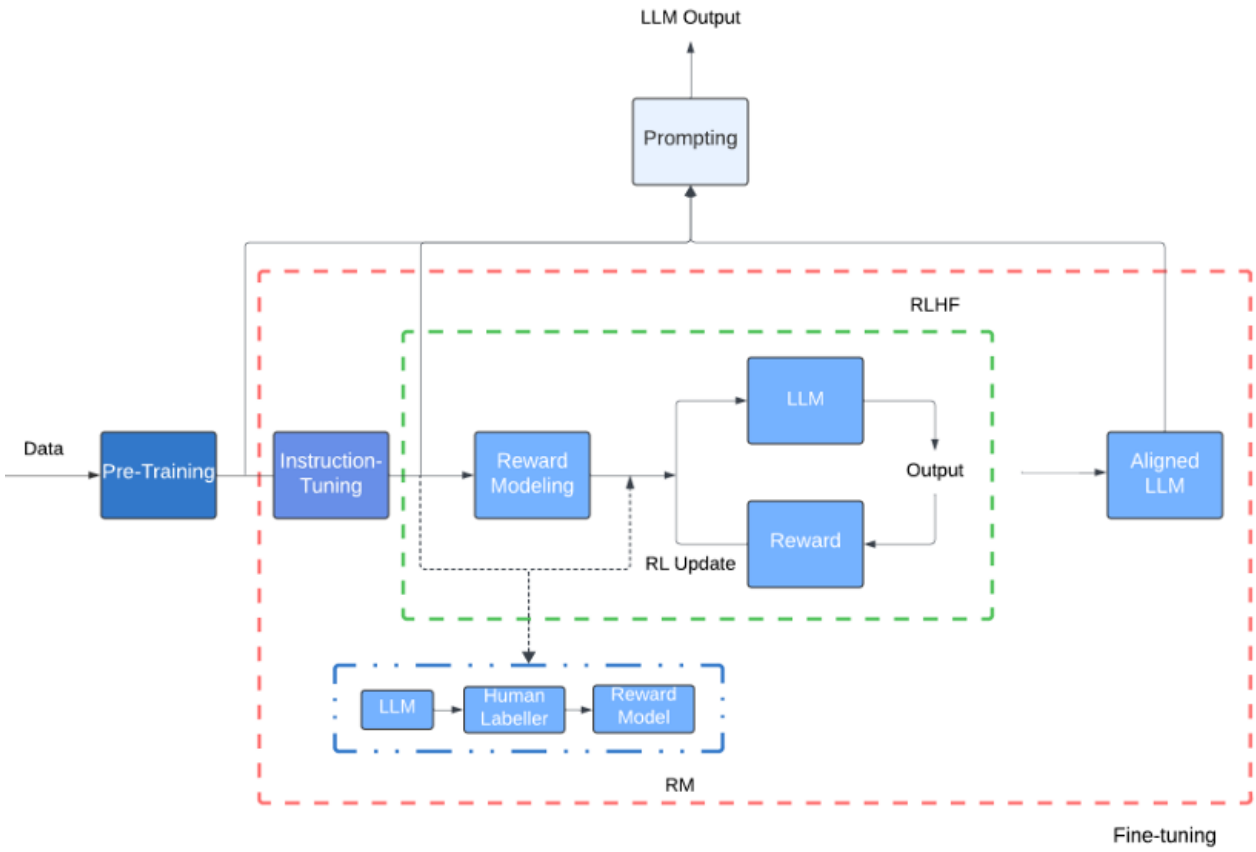

Figure 1: The modern three-stage LLM training pipeline-pre-training, instruction-tuning, and RLHF-adapted from Zhao et al. [20].

Retriever. The Retriever navigates the knowledge graph to surface evidence relevant to the query. Its action set includes (1) *shallow retrieval*, which performs a low-cost local exploration; (2) *deep retrieval*, which performs multi-hop traversal to capture more diverse or distant evidence at higher cost; and (3) *skip retrieval*, which relies on parametric model knowledge when external evidence appears unreliable.

Generator. The Generator produces candidate answers conditioned on retrieved evidence. It may (1) perform a single-pass generation; (2) apply a more robust *self-consistency* strategy, sampling multiple outputs before selecting a consensus; (3) rely solely on parametric knowledge when retrieval is weak; or (4) *abstain* when confidence is low. These options allow the system to vary the strength of generation based on uncertainty.

Verifier. The Verifier checks whether the generated answer is entailed by the retrieved evidence using natural language inference (NLI). It can either run the entailment classifier or skip verification to reduce cost. When run, the Verifier provides an explicit signal about whether the retrieved evidence supports, contradicts, or fails to relate to the generated answer.

Mediator. The Mediator does not execute retrieval, generation, or verification directly. Instead, it orchestrates the agents by selecting an optimal joint action. Its decision is based on a vector of graph-derived reliability signals, including node trustworthiness, evidence-path coherence, subgraph diversity, and local graph connectivity. The Mediator uses these signals to query a CE policy table learned offline, ensuring that the chosen actions are coordinated rather than independent. This coordination is essential: deep retrieval is only useful when paired with verification, and abstention is only rational when both retrieval and generation indicate high risk.

Together, these components enable GRACE to reason adaptively, executing lightweight strategies when evidence is

strong and switching to heavier or safer strategies when uncertainty is detected. The next subsection provides an architectural overview of this coordinated decision process.

3.2 Architecture Overview

Figure 2 illustrates the execution pipeline of GRACE, which transforms the traditional, static RAG workflow into a dynamic, state-conditional decision process. The architecture is organized into four phases:

Reliability Probing and Signal Extraction. To avoid the computational expense of full-scale retrieval and verification for low-quality queries, the system begins with a lightweight probing phase. Upon receiving a user query, the *Entity Linking* module attempts to anchor the question to a specific node v_{start} within KG. Traditional RAG systems often fail silently when the retrieval corpus lacks relevant information. By explicitly probing for v_{start}, we establish an immediate signal of "answerability". If linking succeeds, the system performs a shallow graph traversal to extract topological signals: *Reliability* (PageRank centrality), *Coherence* (consistency of edge relations), and *Diversity* (entropy of node types). These continuous metrics are discretized into a state bin $s \in S$ (e.g., T_{High}, Q_{Low}), compressing the complex graph topology into a tractable state space for decision-making. Crucially, if entity linking fails, the system does not simply error out. Instead, it enters a "No Signal" state ($s_\emptyset$). This architectural choice allows the Mediator to retain control, enabling a graceful degradation to parametric methods rather than a hard crash.

Mediator. At the heart of this architecture lies the Mediator Agent, which functions as the system's policy engine. Unlike rigid pipelines where $Retrieve \rightarrow Generate$ is hard-coded, Mediator queries a learned policy table $\pi(a|s)$ to select a *Joint Action* $\vec{a} = (a_R, a_G, a_V)$. We frame the selection of $\vec{a}$ as a CE problem. This allows the system to balance conflicting objectives: maximizing answer accuracy (utility U_{acc}) while minimizing computational cost (utility U_{cost}) and hallucination risk (utility U_{risk}). For a high-trust state, the Mediator might select a high-fidelity configuration. For a low-trust state, it may rationally opt to skip retrieval and rely on the LLM's internal knowledge, or simply `abstain` if the risk is too high. This dynamic selection prevents the "Potemkin Village" effect, in which models appear correct while relying on irrelevant or misleading evidence [12].

Conditional Execution Loop. Once the Mediator selects a joint action, the system enters an execution loop in which each agent performs its prescribed operation. The loop is conditional rather than sequential: components are activated only when their actions are included in the selected configuration.

Strategic Output Resolution The final architectural safeguard is the gating mechanism. The system distinguishes between "Don't Know" (Abstention) and "Wrong". If the Verifier detects a contradiction (Entailment Score < Threshold), or if the Generator is triggered to refuse to generate by the policy, the system outputs a formalized `Abstain` token. This transforms Abstention from a failure mode into a successful strategic outcome. By explicitly modeling the utility of abstaining ($U_{abstain} > U_{hallucination}$), the architecture incentivizes the system to remain silent rather than confabulate, directly addressing the key reliability deficit of current Generative AI.

3.3 Correlated Equilibrium Policy

GRACE coordinates the Retriever, Generator, and Verifier through a centralized policy derived CE. This section formalizes the decision process and describes how graph-derived signals shape action selection.

Game Structure. We model the interaction between functional modules as a cooperative game defined by the tuple $\langle N, S, A, U \rangle$:

Players. The set of players is N = {Retrieval, Generation, Verification}. These modules do not compete; they act cooperatively to maximize the shared utility U. Each module $i \in N$ selects an action a_i from its action set: $a_R \in A_R$, $a_G \in A_G$, $a_V \in A_V$

State Space. The state $s \in S$ is a discrete encoding of the graph context surrounding the query entity. The discretization maps continuous reliability signals into a low-dimensional representation, making the game tractable for learning and inference.

Joint Action. The global action is the Cartesian product

$$a = (a_R, a_G, a_V) \in A_R \times A_G \times A_V.$$

Examples include (`retrieve_deep`, `generate_consistency`, `verify_nli`) or (`skip_retrieval`, `generate_parametric`, `skip_verify`). The expressiveness of A enables GRACE to balance cost, accuracy, and risk.

Utility Function. For any state-action pair (s, a), the utility is defined as

$$U(s, a) = \mathbb{I}_{\text{correct}} \cdot R_{\text{correct}} + \mathbb{I}_{\text{abstain}} \cdot R_{\text{safe}} - \sum_{i \in N} \text{Cost}(a_i).$$

Here, as we went through several rounds of calibration, we set correct answers receive high reward; safe abstentions receive smaller positive reward; and computational cost is penalized additively across modules. Incorrect answers incur zero positive reward and incur the full cost of the executed

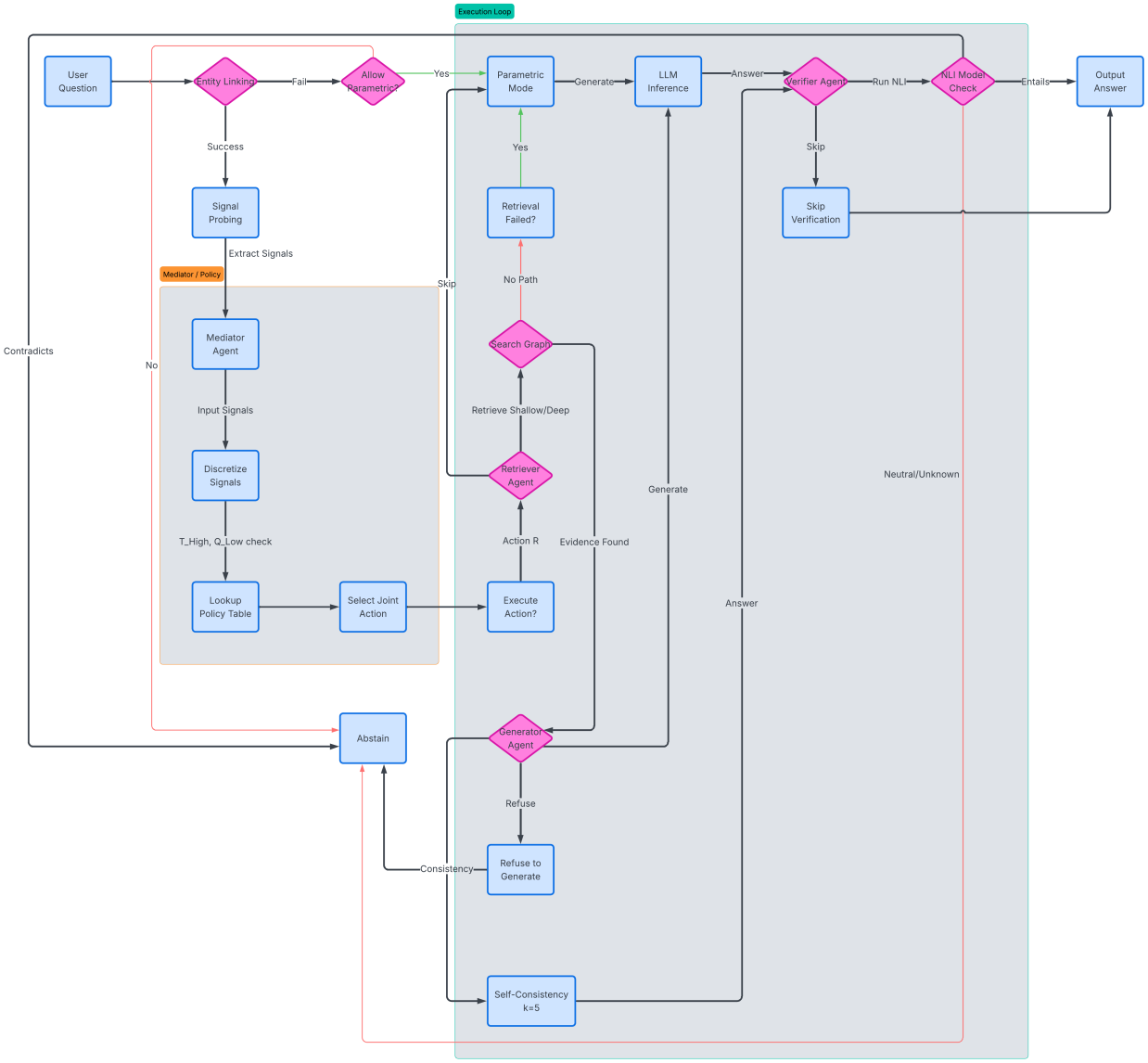

Figure 2: System architecture of GRACE. Given a user query, the system performs entity linking and graph probing to extract structural reliability signals. These signals are discretized and passed to a central Mediator, which consults a CE policy to select a coordinated joint action for the Retriever, Generator, and Verifier. The selected actions determine how evidence is gathered, how answers are generated, and whether verification or abstention is performed.

action sequence. This design ensures that unnecessary computation (e.g., deep retrieval when signals are weak) is disfavored. We decided to design as such to ensure that abstention is permitted but not dominant: if R_{safe} were set too high or R_{correct} too low, the optimal policy would collapse into frequent abstention even in states where reliable reasoning is possible, reducing the usefulness of the system.

Value Estimation. During simulation, we approximate the state-action value function

$$Q(s, a) = \mathbb{E}[U(s, a)]$$

empirically by repeatedly executing the system under randomized policies and recording observed utilities. This yields

sample-based estimates $\widehat{Q}(s, a)$ used during policy derivation.

Stochastic Policy via Quantal Response. Instead of selecting the single highest-valued action, we compute a stochastic policy

$$\pi(a \mid s) = \frac{\exp\left(\widehat{Q}(s, a)/\tau\right)}{\sum_{a' \in A} \exp\left(\widehat{Q}(s, a')/\tau\right)},$$

where $\tau > 0$ is a temperature parameter. Low τ yields near-greedy selection; high τ induces broader exploration. This formulation corresponds to a *Quantal Response Equilibrium* (QRE) [13], which models players selecting higher-utility actions with higher probability while tolerating bounded noise. QRE is well-suited for GRACE because the estimated $\widehat{Q}$ values contain stochastic variability due to retrieval noise and generation variance.

3.4 Signals and States

To support efficient policy lookup, high-dimensional graph signals are compressed into a discrete state space.

Raw Signals. Given the entity-linked node v_{start}, we extract five continuous metrics from its local neighborhood:

(1) Reliability ρ: mean PageRank score of reachable nodes.

(2) Degree δ: local connectivity (information density).

(3) Path length λ: estimated mean hop-distance to plausible answers.

(4) Coherence γ: relational consistency among traversed paths.

(5) Diversity H: entropy over encountered entity types.

These signals capture complementary aspects of the graph structure: trustworthiness of sources, accessibility of evidence, and quality of relational information.

Meta-Signal Aggregation. To avoid an excessively large state space, we combine the raw signals into two meta-dimensions:

$$T = \text{Bin}(\rho) + \text{Bin}(\delta), \quad Q = \text{Bin}(\lambda) + \text{Bin}(\gamma) + \text{Bin}(H),$$

where $\text{Bin}(\cdot)$ applies quantile-based discretization. T reflects the reliability of the local graph region; Q reflects the quality and coherence of accessible evidence.

Final State Representation. Each query is assigned a state

$$s = (T, Q),$$

with $T, Q \in \{\text{Low}, \text{Mid}, \text{High}\}$, yielding $|S| = 9$ states. This compact representation preserves the most informative structural variability while enabling rapid inference and stable policy learning.

4 GRACE IMPLEMENTATION

4.1 Experimental Setup

We evaluate GRACE on the HotpotQA distractor split [18], a setting that requires multi-hop reasoning in the presence of intentionally misleading evidence. This environment is well suited for assessing whether GRACE can detect unreliable conditions and adjust its retrieval and generation procedures accordingly. To supply the structural signals required by the Mediator, we construct a knowledge graph G from entity mentions and supporting sentences, treating entities as nodes and semantic relations as edges. This graph-based representation exposes topological cues—reliability, coherence, and diversity—that would be inaccessible in a purely text-based retrieval pipeline.

To learn the CE policy, we generate simulated trajectories on a calibration subset ($N_{\text{calib}} = 100$). This allows the system to observe the utility consequences of different joint actions across a variety of graph-derived states and to estimate $\widehat{Q}(s, a)$ values. A separate held-out evaluation split ensures that the learned policy generalizes rather than overfitting to the calibration dynamics.

Our evaluation of GRACE is structured around a question: *Does coordinated, state-dependent decision-making yield a more reliable and economically efficient reasoning pipeline than static alternatives?* To answer this, we evaluate GRACE against three baselines, each representing a distinct philosophy of QA system design. For clarity, the setup is summarized as follows:

(1) **Baseline (No Retrieval).** This model answers using only its internal weights, providing the lower bound on external grounding. It reveals whether retrieval is necessary for the dataset and quantifies the hallucination risk inherent in closed-book generation.

(2) **Baseline + Standard GraphRAG (Deterministic Retrieval).** This pipeline always executes the same shallow graph retrieval and always generates an answer regardless of evidence quality. It reflects prevailing industry practice and exposes a key failure mode: retrieving irrelevant subgraphs can degrade accuracy by contaminating the generator's belief state.

(3) **Baseline + CE (Token-Level Refusal).** This heuristic model abstains when the generator's token-level confidence falls below a fixed threshold. It serves as

the strongest simple alternative to GRACE by asking whether internal model confidence alone is sufficient to achieve reliable refusal. If this baseline matched GRACE's reliability or utility, graph-derived structural signals and game-theoretic coordination would offer limited additional value.

Together, these baselines allow us to isolate the contribution of GRACE's state-conditioned, utility-driven orchestration rather than any advantage conferred by model scale or internal memorization.

All model components are intentionally lightweight so that improvements can be attributed to the framework rather than scale. We employ FLAN-T5-Base (250M parameters) [2] as the generator, relying on temperature sampling for self-consistency and greedy decoding for standard outputs. This choice ensures that gains in reliability cannot be attributed to memorization or parametric knowledge alone. Verification is handled by a RoBERTa-Large NLI classifier trained on standard entailment corpora, selected for its speed and stability relative to generative judges. Retrieval is implemented as a bounded-width heuristic search in NetworkX so that graph-signal extraction remains tractable. All experiments are conducted in a single-GPU environment with CUDA acceleration, reflecting realistic constraints for mid-scale research systems.

The evaluation emphasizes reliability and decision quality rather than raw accuracy. Correctness is measured by a BERTScore F1 threshold of 0.9, ensuring semantic equivalence rather than superficial lexical overlap. Abstention rate quantifies the system's willingness to invoke the safe fallback mechanism central to GRACE's design. Effective reliability reports accuracy conditional on answered queries, providing insight into whether abstention eliminates weak predictions or merely hides errors. Finally, net utility reflects the game-theoretic payoff achieved under the learned policy and serves as the primary indicator of whether GRACE achieves a superior balance between correctness, safety, and computational cost.

4.2 Result and Discussion

Figure 3 reports the performance of GRACE and the three baseline systems across overall accuracy, abstention rate, effective reliability, and net utility. The results reveal a systematic divergence between recall-oriented metrics (accuracy) and precision-oriented metrics (effective reliability), highlighting how GRACE's adaptive, game-theoretic design reshapes the balance between correctness and coverage.

Contrary to the initial expectation that a CE policy would maximize expected utility by balancing accuracy and cost, the empirical results show that GRACE behaves as an *too conservative* system. It consistently prioritizes safety over coverage, yielding markedly lower overall accuracy than the baselines but substantially reducing erroneous predictions. This trade-off is visible in the strong decoupling between recall and precision: GRACE answers far fewer questions, but the answers it does provide remain reasonably reliable despite noisy retrieval conditions.

A key finding emerges when comparing the Parametric Baseline and the deterministic GraphRAG pipeline. The closed-book parametric model achieves the highest accuracy (30.0%), while GraphRAG drops to 20.0%. This indicates that retrieval was detrimental for this subset of HotpotQA. Many retrieved subgraphs contain distractors or weakly related nodes. The generator, conditioned on these irrelevant paths, exhibits "hallucination by contamination", performing worse than when relying solely on internal knowledge. GRACE detects these low-trust situations via its graph signals, but because its action space cannot repair poor retrieval, abstention becomes the rational choice.

The CE baseline achieves the highest effective reliability (80.0% precision), outperforming GRACE (50.0%). This suggests that token-level confidence in FLAN-T5 currently provides a stronger predictor of correctness than the graph-derived signals used here. While the heuristic abstains only in low-confidence cases, GRACE abstains whenever the graph signals suggest uncertainty—which, due to signal compression and noise, happens far more often. Thus, GRACE's limitations stem not from the CE framework but from the weak discriminative power of upstream graph signals.

These quantitative observations motivate a deeper question: Why did the game-theoretic agent converge on such extreme passivity?

Signal Compression. PageRank normalization forces question nodes to dominate, flattening differences among entity nodes and preventing meaningful gradients in reliability. Without discriminative signals, the Mediator is effectively flying blind.

Retrieval Noise. The "distractor effect" in dense graph regions is severe, frequently misaligning the system's trust perception.

Miscalibrated Incentives.: The utility function is overly risk-averse. The combination of high penalties for error and high costs for verification makes abstention the statistically dominant strategy in all but the clearest states.

To visualize how these structural factors shaped the agent's decision-making process, Figure 4 maps the learned CE policy across the discretized state space.

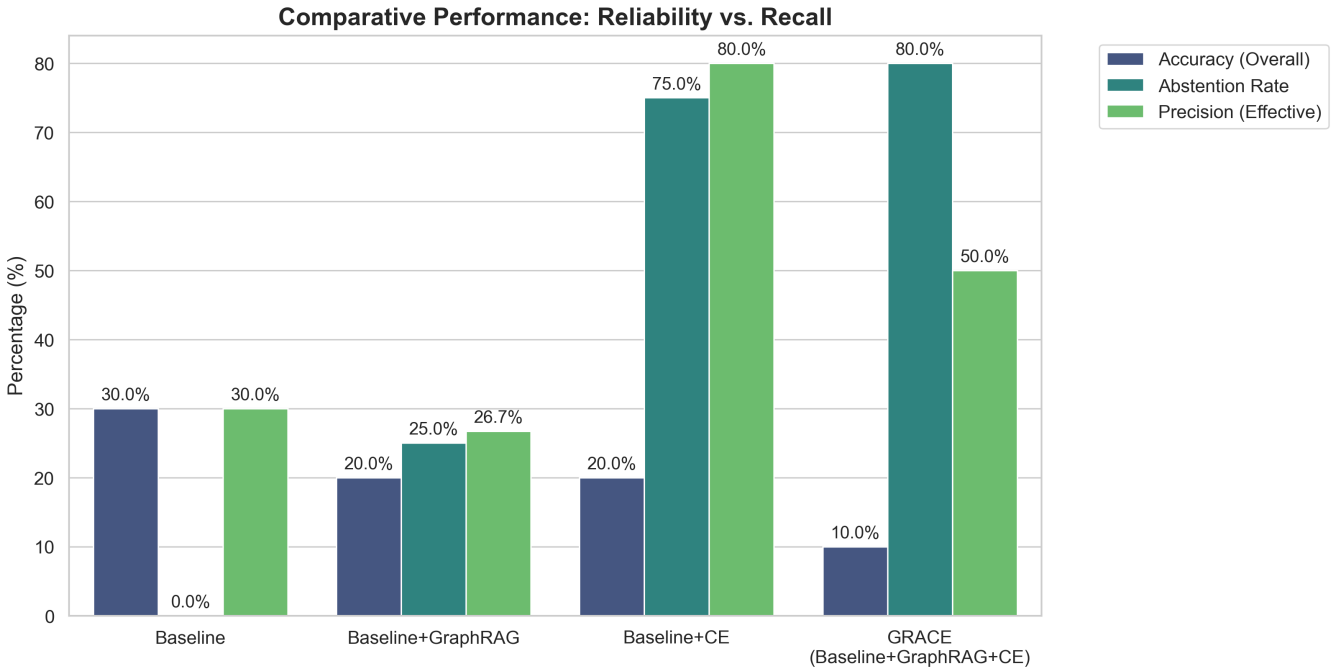

Figure 3: Comparison of GRACE with three baselines across accuracy, abstention rate, effective reliability, and net utility. GRACE exhibits strong safety behavior driven by risk-aware utility shaping but suffers from high abstention caused by noisy graph signals and retrieval bottlenecks.

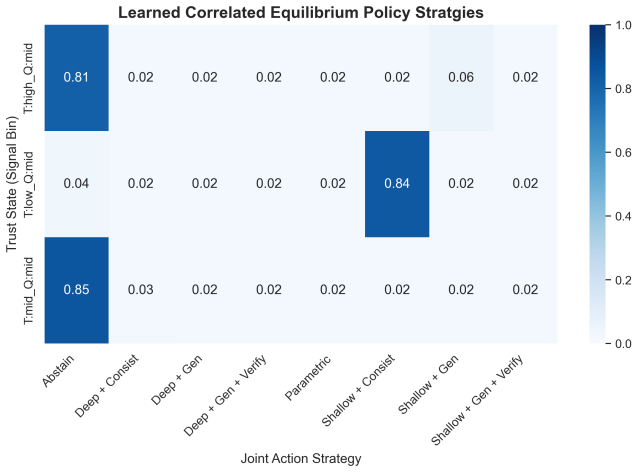

Figure 4: Learned CE policy across discretized graph-signal states. Each row corresponds to a (T, Q) state, and each column to a joint action. The policy exhibits extreme conservatism, selecting refusal-heavy strategies even when structural signals appear favorable.

The heatmap confirms the impact of the structural bottlenecks identified above. The policy is highly deterministic and polarized. In mid- and high-trust states, the dominant action is *Retrieve Shallow + Refuse to Generate* ($P \approx 0.79$). Despite "High Trust" topology, the Mediator learned that the unobserved retrieval noise makes answering net-negative. Conversely, in low-trust states (T:*low*, Q:*mid*), the policy flips to *Generate with Consistency* ($P \approx 0.84$). This aligns with our signal analysis: sparse graphs contain less noise, so the signal-to-noise ratio is paradoxically better in "Low Trust" states, justifying the high cost of verification.

This polarization is further explained by the incentives shown in Figure 5. The utility distribution reveals an "Incentive Cliff". In dense regions, the variance of answering is dangerous (rewards of +20 vs penalties of −40), pushing the agent to the safe harbor of abstention (clustered around +10). In sparse regions, the variance expands significantly, with successful Consistency checks yielding up to +114.0 (as seen in Case 1 below).

To ground these statistics in system behavior, we analyze four decisions observed in the evaluation logs:

Case 1: Success in Low Trust. For a question asking, *"Which performance act has a higher instrument to person ratio...?"*, with the correct answer being "Badly Drawn Boy", the system detects a sparse graph state (T:*low*, Q:*mid*). Despite the low trust signal, the retrieval finds a highly informative path

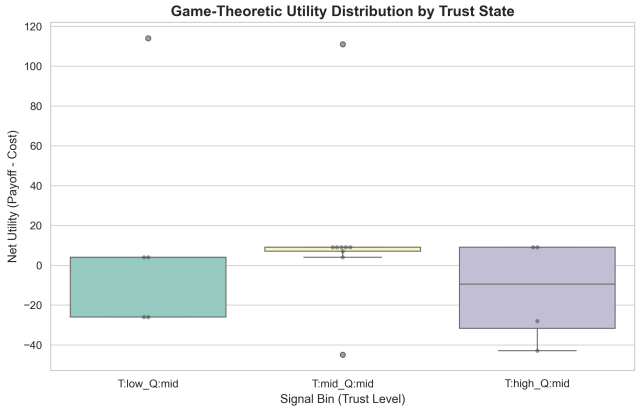

Figure 5: Distribution of net utility across discretized trust states under the learned CE policy. High-trust regions show tightly clustered abstention payoffs, while low-trust regions exhibit high-variance outcomes associated with consistency-based answering.

explicitly defining Badly Drawn Boy as "an English indie singer-songwriter and multi-instrumentalist". Identifying this high-quality semantic match, the system commits to *Generate Consistency*. The risk pays off: the explicit definition allows the model to correctly infer a high ratio, achieving a net utility of +114.0. This confirms GRACE can be aggressive when it detects high-information paths in low-noise environments.

Case 2: Failure in High Trust. For a question asks for the founding year of VCU (1838). The central entities trigger a (T:*high*, Q:*mid*) signal, often implying safety. However, the retrieval fails to find the founding node and instead pulls a semantic distractor: the "2011–12 VCU Rams men's basketball team". Misled by the "High Trust" topology signal, the policy skips verification to save cost. The Generator, conditioned on the only year present in the context ("2011"), hallucinates this as the founding date. This results in a penalty of $-\mathbf{43.0}$, illustrating how topological trust signals can fail when semantic retrieval is noisy.

Case 3: Correct Abstention. For a question asking about a "Mexican Formula One driver". The retrieval agent completely fails, returning unrelated paths about Pakistani cricketers ("Aamer Iqbal"). The baseline model, forced to answer, claims a hallucinated name ("antonio saad"). GRACE, however, runs a consistency check. Finding no semantic link between "Cricket" evidence and "Formula One" questions, the generator produces inconsistent or refusal outputs across the $k = 5$ samples. The Mediator correctly interprets this dissonance and abstains. While the reward is small (+4.0), the system successfully shields itself from a hallucination event caused by total retrieval failure.

Case 4: False Negative. For a simple binary question: *"Are both Dictyosperma, and Huernia described as a genus?"*, the retrieved evidence actually contains the answer ("The genus is named..."), but the path coherence signal is low (Q:*low*). The policy, conditioned to be ultra-conservative by the high penalties in Case 2, interprets this structural noise as a risk of hallucination and refuses to generate. The Baseline answers correctly ("yes"), leaving GRACE with an opportunity cost of −21.0. This highlights the trade-off of the learned refusal strategy: valid answers are sacrificed to maintain high aggregate precision.

GRACE demonstrates the mechanical capacity for dynamic coordination—switching between aggression (Case 1) and defense (Case 3) based on state. However, its practical performance is currently limited by the fidelity of its signals and the severity of its reward structure, leading to the "honest but too conservative" behavior pattern.

5 RELATED WORK

5.1 Uncertainty Estimation and Abstention

Selective prediction aims to improve reliability by allowing models to abstain under uncertainty. Early work formalized selective classification as an optimization problem [4]. Subsequent extensions in NLP [3] show that abstention can substantially improve precision in generative tasks. Language-model uncertainty has been studied through token-level confidence [5], calibration [10], semantic uncertainty [8], and multi-path aggregation such as self-consistency [16].

External verifiers based on NLI have also been used to detect factual inconsistencies [14, 15]. However, these approaches typically apply verification uniformly, without accounting for retrieval quality, computational cost, or the differential risk structure across queries.

GRACE differs in treating abstention as a strategic choice derived from expected utility, rather than a threshold on model confidence. It further integrates retrieval uncertainty via graph-structural signals, enabling abstention or verification to be invoked selectively.

5.2 Game-Theoretic Control

With the rise of tool-augmented LLMs, several works have explored structured decision-making for multi-step reasoning. Approaches such as ReAct [19], reasoning executives [17], and decomposed tool-use strategies [7] show that explicit decision policies can improve control, though they are typically heuristic or prompt-based.

Game-theoretic formalisms such as CE [1] and QRE [13] offer principled coordination under uncertainty and bounded rationality. While widely used in multi-agent systems and behavioral modeling, they have rarely been applied to retrieval-augmented QA pipelines.

GRACE applies CE as a mediator that jointly coordinates Retrieval, Generation, and Verification as cooperative decision-makers with associated costs. This places RAG workflows on a principled decision-theoretic foundation rather than relying on ad-hoc triggers or static pipelines.

6 CONCLUSION

In this work, we introduced GRACE, a multi-agent framework that orchestrates RAG using CE. By modeling the interaction between Retrieval, Generation, and Verification as a cooperative game, we sought to replace static heuristic pipelines with dynamic, risk-aware decision-making.

Our experimental evaluation reveals that GRACE successfully internalizes the logic of strategic coordination. The system learned to act as a rigorous "Safety Filter", achieving high effective reliability by screening out hallucinations and noisy retrieval artifacts. The qualitative analysis confirms that the agent is capable of sophisticated behaviors, switching between aggressive consistency-checking in low-noise environments and defensive abstention in high-noise states.

However, the system's performance is currently constrained by the "Incentive Cliff" of its reward structure and the compression of its graph signals. Faced with high penalties for error and low-fidelity structural signals, the CE policy converged on an *too conservative* strategy, prioritizing safety over coverage to an excessive degree. While this behavior is rational under the defined utility function, it limits the system's practical recall.

These findings point to a clear path for future research. The coordination mechanism itself is sound; the bottleneck lies in the system's perception. Future iterations must focus on deriving higher-fidelity graph signals—perhaps through hierarchical PageRank or semantic density metrics—that can distinguish true opportunities from distractors. Additionally, moving from binary success/failure rewards to graded utility functions could encourage the agent to explore the middle ground between silence and high-stakes gambling. Ultimately, GRACE demonstrates that game-theoretic control is a viable paradigm for robust AI systems, provided the agents are equipped with the sensory precision to navigate the risks they compute.

REFERENCES

[1] Robert J. Aumann. 1974. Subjectivity and Correlation in Randomized Strategies. *Journal of Mathematical Economics* (1974).

[2] Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Eric Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zihang Dai, Ozlem Uzuner, Jason Wei, Tom Brown, Denny Zhou, Quoc V. Le, Jeff Dean, Sharan Narang, Prateek Shyam, Maarten Bosma, Yanping Zhao, Zheng Chen, and Adam Roberts. 2022. Scaling Instruction-Finetuned Language Models. *arXiv preprint arXiv:2210.11416* (2022).

[3] Ohad Elyashiv, Ben Feinstein, and Ran El-Yaniv. 2023. Selective Prediction for Natural Language Processing. *Transactions of the Association for Computational Linguistics* (2023).

[4] Yoad Geifman and Ran El-Yaniv. 2017. Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems*.

[5] Saurav Kadavath et al. 2022. Language Models (Mostly) Know What They Know. *arXiv preprint arXiv:2207.05221* (2022).

[6] Adam Tauman Kalai, Ofir Nachum, Santosh S. Vempala, and Edwin Zhang. 2025. Why Language Models Hallucinate. *arXiv preprint* (2025). arXiv:2509.04664 [cs.LG] https://arxiv.org/abs/2509.04664

[7] Tushar Khot et al. 2023. Decomposed Prompting for Reinforcement Learning and Tool Use. *arXiv preprint arXiv:2302.13189* (2023).

[8] Lukas Kuhn et al. 2023. Semantic Uncertainty in Large Language Models. *arXiv preprint arXiv:2302.09664* (2023).

[9] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems (NeurIPS 2020)*. Curran Associates, Inc. https://arxiv.org/abs/2005.11401

[10] Stephanie Lin et al. 2022. Teaching Models to Refuse Unknown Questions. In *Proceedings of EMNLP*.

[11] Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Percy Liang, and Tatsunori Hashimoto. 2023. Lost in the Middle: How Language Models Use Long Contexts. *arXiv preprint arXiv:2307.03172* (2023).

[12] Emmanuelle Mathieu, Nitish Joshi, and Alexander M. Rush. 2020. Potemkin Models: Weaknesses of Learned Evaluation Metrics. In *EMNLP*.

[13] Richard D. McKelvey and Thomas R. Palfrey. 1995. Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior* 10, 1 (1995), 6–38.

[14] Yixin Nie, Adina Williams, Emily Dinan, Dhruv Batra, and Douwe Kiela. 2019. Dissent: Sentence-level Fact Checking with Natural Language Inference. In *Proceedings of ACL*.

[15] Tal Schuster, Adam Fisch, and Regina Barzilay. 2022. Get Your Facts Right: Factuality Evaluation for Text Generation with NLI. In *Proceedings of ACL*.

[16] Xuezhi Wang et al. 2022. Self-Consistency Improves Chain of Thought Reasoning. *arXiv preprint arXiv:2203.11171* (2022).

[17] Zhaofeng Wu et al. 2023. Reasoning Executives for Large Language Models. *arXiv preprint arXiv:2305.14955* (2023).

[18] Zhiguo Yang, Peng Qi, Saizheng Zhang, Jason Bolton, Christopher D. Manning, William Cohen, Chris Dyer, and Xiaodong Lin. 2018. HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2369–2380.

[19] Shunyu Yao et al. 2022. ReAct: Synergizing Reasoning and Acting in Language Models. *arXiv preprint arXiv:2210.03629* (2022).

[20] Wayne Xin Zhao, Jingyuan Jiang, et al. 2023. A Comprehensive Overview of Large Language Models. *arXiv preprint arXiv:2307.06435* (2023).