

# Adaptive Root Cause Localization for Microservice Systems with Multi-Agent Recursion-of-Thought

Lingzhe Zhang, Tong Jia\*, Kangjin Wang, Weijie Hong, Chiming Duan,  
Minghua He, and Ying Li\*, *Member, IEEE*

**Abstract**—As contemporary microservice systems become increasingly popular and complex—often comprising hundreds or even thousands of fine-grained, interdependent subsystems—they are facing more frequent failures. Ensuring system reliability thus demands accurate root cause localization. While traces and metrics have proven to be effective data sources for this task, existing methods either heavily rely on pre-defined schemas, which struggle to adapt to evolving operational contexts, or lack interpretability in their reasoning process, thereby leaving Site Reliability Engineers (SREs) confused. In this paper, we conduct a comprehensive study on how SREs localize the root cause of failures, drawing insights from multiple professional SREs across different organizations. Our investigation reveals that human root cause analysis exhibits three key characteristics: recursiveness, multi-dimensional expansion, and cross-modal reasoning. Motivated by these findings, we introduce RCLAgent, an adaptive root cause localization method for microservice systems that leverages a multi-agent recursion-of-thought framework. RCLAgent employs a novel recursion-of-thought strategy to guide the LLM’s reasoning process, effectively integrating data from multiple agents and tool-assisted analysis to accurately pinpoint the root cause. Experimental evaluations on various public datasets demonstrate that RCLAgent achieves superior performance by localizing the root cause using only a single request—outperforming state-of-the-art methods that depend on aggregating multiple requests. These results underscore the effectiveness of RCLAgent in enhancing the efficiency and precision of root cause localization in complex microservice environments.

**Index Terms**—Root Cause Localization, Trace, Multi-Agent, Recursion-of-Thought.

## I. INTRODUCTION

Modern microservice systems have become increasingly complex due to dynamic interactions and evolving runtime environments [1]–[3]. These systems often consist of hundreds or even thousands of fine-grained, interdependent subsystems, where issues in any one component can easily lead to performance problems at the top level [4]–[8]. Therefore, to ensure system reliability, it is crucial to localize the root cause of these issues in a timely manner [9], [10].

However, localizing the root cause of these issues is challenging due to the intricate dependencies between subsystems

Lingzhe Zhang, Tong Jia, Weijie Hong, Chiming Duan, Minghua He, and Ying Li are with Peking University, Beijing, China.

Email: {zhang.lingzhe, hongwj, duanchiming, hemh2120}@stu.pku.edu.cn and {jia.tong, li.ying}@pku.edu.cn

Kangjin Wang is with Alibaba Group, China, e-mail: (kangjin.wkj@alibaba-inc.com)

\* Corresponding author: Tong Jia, e-mail: (jia.tong@pku.edu.cn); Ying Li, e-mail: (li.ying@pku.edu.cn)

Manuscript received May 28, 2025; revised xxx.

within microservice systems [11]–[18]. Each request typically follows complex invocation chains, with each chain involving multiple components—such as services, service instances, and hosts—and the interactions between them, including service calls, database queries, and other inter-component communications. Moreover, the dynamic nature of these interactions, combined with the heterogeneity of system components, makes pinpointing the root cause a highly non-trivial task.

To enable fine-grained root cause localization of these issues, extensive work leverages both trace and metrics data from software systems. Metrics data provides aggregate insights into system performance (e.g., response times, throughput, and resource utilization), as exemplified by approaches such as Microscope [19], which constructs causality graphs and employs a depth-first search strategy to identify front-end anomalies; CIRCA [20], which builds a causal Bayesian network using regression-based hypothesis testing and descendant adjustment to infer problematic components; and RUN [21], which employs time series forecasting for neural Granger causal discovery and integrates a personalized PageRank algorithm to efficiently recommend the top-k root causes. Meanwhile, trace data offers a more granular view by tracking the execution paths of individual requests and inter-component interactions, as demonstrated by MicroRank [22], which designs a trace coverage tree to capture dependencies between requests and service instances and applies the PageRank algorithm to score potential root causes; TraceRank [23], which combines spectrum analysis with a PageRank-based random walk to pinpoint abnormal services; CRISP [24], which performs critical path analysis tailored to drill down latency issues; and TraceConstruct [25], which uses sequence representations along with contrast sequential pattern mining and spectrum analysis to efficiently localize multi-dimensional root causes. Some approaches integrate multimodal data, particularly LLM-based methods. For instance, mABC [26] introduces a multi-agent, blockchain-inspired collaboration framework where multiple LLM-based agents follow a structured workflow and coordinate through blockchain-inspired voting mechanisms. RCAgent [27] leverages log and code data in a tool-augmented LLM framework to perform root cause analysis tasks such as predicting root causes, identifying solutions, gathering evidence, and determining responsibilities.

Although these root cause localization methods have demonstrated promising outcomes, they still face the following practical challenges when applied to real-world microservice systems:

- **Heavy Reliance on Pre-defined Schemas.** Most ap-

proaches depend on pre-defined service causal graphs, fixed statistical models of fault-symptom relationships, or implicitly learned associations from training data—essentially relying on static, one-time analyses. While such schemas provide a baseline capability for root cause localization, they also impose significant limitations on generalizability. The effectiveness of these models is strictly tied to their underlying assumptions; any changes in service dependencies or predefined anomaly models can severely compromise performance. Consequently, these rigid methods struggle to adapt to evolving operational contexts.

- **Lack of Interpretability in the Reasoning Process.**

Although some models use deep learning to dynamically capture service dependencies and their correlation with root causes, they often suffer from a lack of transparency compared to graph-based approaches. Given a formatted input, these models can produce a list of potential root causes but typically fail to explain how or why these conclusions are reached. In contrast, an interpretable reasoning process is crucial for Site Reliability Engineers (SREs), as it enables rapid verification of the root cause and facilitates timely remediation.

Recognizing this gap, we first conduct an empirical study to investigate how SREs localize the root causes, drawing insights from multiple professional SREs across different organizations. This study reveals three key characteristics of manual root cause analysis: recursiveness (when a deeper-level root cause is identified, SREs iteratively refine their analysis by examining lower-layer manifestations), multi-dimensional expansion (broadening the search across different dimensions, such as pods, services, and infrastructure, to account for all potential root causes), and cross-modal reasoning (once a potential root cause is identified through trace data, it is validated by analyzing fluctuations in relevant metrics).

Building on these insights, we introduce **RCLAgent**<sup>1</sup>, an adaptive root cause localization method for microservice systems with multi-agent recursion-of-thought. RCLAgent comprises two types of agents: data agents, which integrate trace and metrics data, and thought agents, which mainly employ recursion-of-thought to iteratively infer and refine the root cause. These agents are coordinated by a central coordinator, which orchestrates the process through three key phases: initial reasoning, critical reflection, and final review.

Our experiments demonstrate that RCLAgent significantly outperforms existing state-of-the-art root cause localization methods. The evaluation results show that RCLAgent delivers superior performance by analyzing a single request, outperforming alternative methods that require multiple requests for analysis. Notably, RCLAgent's recall@1 even surpasses the recall@10 of competing approaches. Furthermore, when a simple majority voting strategy is employed, RCLAgent's group ranking MRR exhibits a substantial improvement, surpassing the second-best method by an average of 32.53%. In summary, the key contributions of this work are as follows:

- We conduct a comprehensive study on how SREs localize the root cause of an abnormal request. Our findings reveal that human root cause analysis exhibits three key characteristics: recursiveness, multi-dimensional expansion, and cross-modal reasoning.
- Inspired by these findings, we propose RCLAgent, an adaptive root cause localization method for microservice systems, leveraging a multi-agent recursion-of-thought framework.
- We evaluate RCLAgent on six datasets, demonstrating its effectiveness. Experimental results show that RCLAgent achieves superior performance by analyzing a single request, surpassing state-of-the-art methods that rely on multiple requests for analysis.

## II. BACKGROUND

In this section, we present the essential background of this paper, including the formal definition of the root cause localization problem, an overview of traditional root cause localization methods, and an introduction to distributed tracing as the primary data source for root cause localization.

### A. Root Cause Localization

Failure diagnosis in distributed systems is generally divided into two categories: failure category classification and root cause localization. The former determines the failure type, such as CPU or memory anomalies, while the latter pinpoints the specific node, service, or pod responsible for the issue. Root cause localization is particularly critical for minimizing downtime and ensuring system stability.

Traditional root cause localization methods typically analyze a collection of anomalous requests within a predefined time window. These methods construct a dependency graph where nodes represent system components, and edges capture causal relationships inferred from request traces or system logs. Graph-based algorithms, such as PageRank, are then applied to rank components based on their likelihood of being the root cause.

$$C^* = C^* = \arg \max_{C \in \mathcal{C}} s(C, G, M) \quad (1)$$

Formally, given a set of anomalous requests  $R = \{r_1, r_2, \dots, r_n\}$ , a component graph  $G = (C, E)$  is constructed, where  $C$  is the set of components, and  $E$  represents their dependencies. The root cause component  $C^*$  is identified as Equation 1, where  $s(C, G, M)$  represents the computed score incorporating graph topology and observed anomalies.

### B. Distributed Tracing

To support fine-grained fault diagnosis in software systems, distributed tracing has been widely adopted in industrial environments, becoming an integral part of modern software infrastructures [28]–[30]. A distributed trace provides a detailed execution record of a request as it propagates through the system, capturing timing, dependencies, and performance characteristics. Each trace consists of multiple structured log

<sup>1</sup>Code Repository: <https://github.com/LLMLog/RCLAgent>

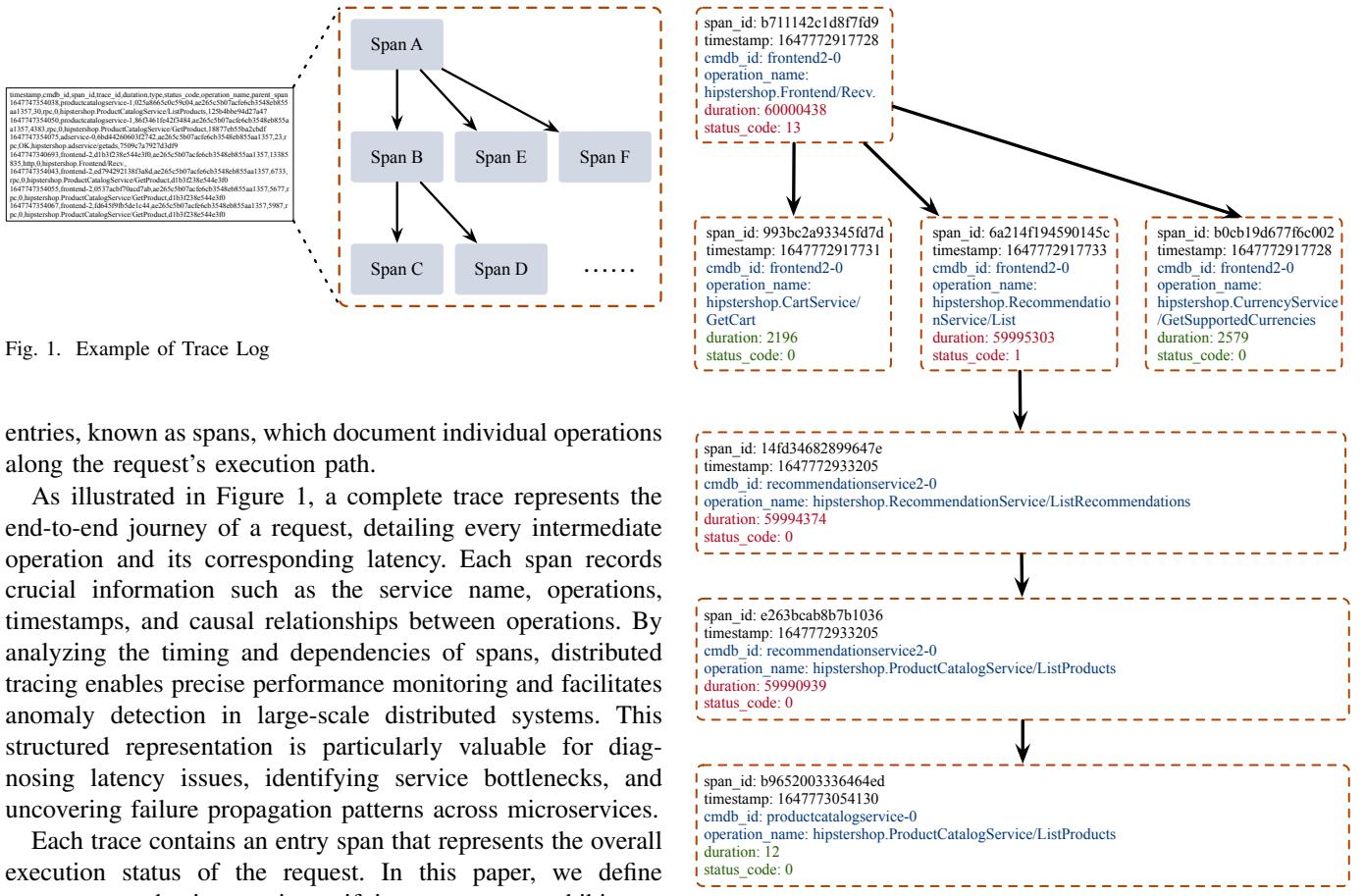


Fig. 1. Example of Trace Log

entries, known as spans, which document individual operations along the request's execution path.

As illustrated in Figure 1, a complete trace represents the end-to-end journey of a request, detailing every intermediate operation and its corresponding latency. Each span records crucial information such as the service name, operations, timestamps, and causal relationships between operations. By analyzing the timing and dependencies of spans, distributed tracing enables precise performance monitoring and facilitates anomaly detection in large-scale distributed systems. This structured representation is particularly valuable for diagnosing latency issues, identifying service bottlenecks, and uncovering failure propagation patterns across microservices.

Each trace contains an entry span that represents the overall execution status of the request. In this paper, we define a request as having an issue if its entry span exhibits an excessively high execution latency—specifically, exceeding 100 times the normal average latency. In such cases, root cause localization is necessary to identify the underlying source of the anomaly.

### III. EMPIRICAL STUDY

In this section, we conduct a comprehensive study on how SREs localize the root cause of an abnormal request, with insights gathered from multiple professional SREs across different organizations. To systematically analyze this process, we investigate the following three research questions:

- **RQ1:** How do SREs initially narrow down the search space for the root cause?
- **RQ2:** How do they systematically expand the search scope to encompass all potential root causes?
- **RQ3:** How do they differentiate the actual root cause from other potential but non-causal anomalies?

#### A. Recursiveness

Deep-seated issues often manifest as a request that is exceptionally slow or even fails. To narrow down the search space for the root cause, human analysts begin with the problematic request and examine its trace data to identify the invoked services and executed operations. If anomalies are detected in these services, they further investigate the downstream services they call, recursively refining the search until the true root cause is isolated.

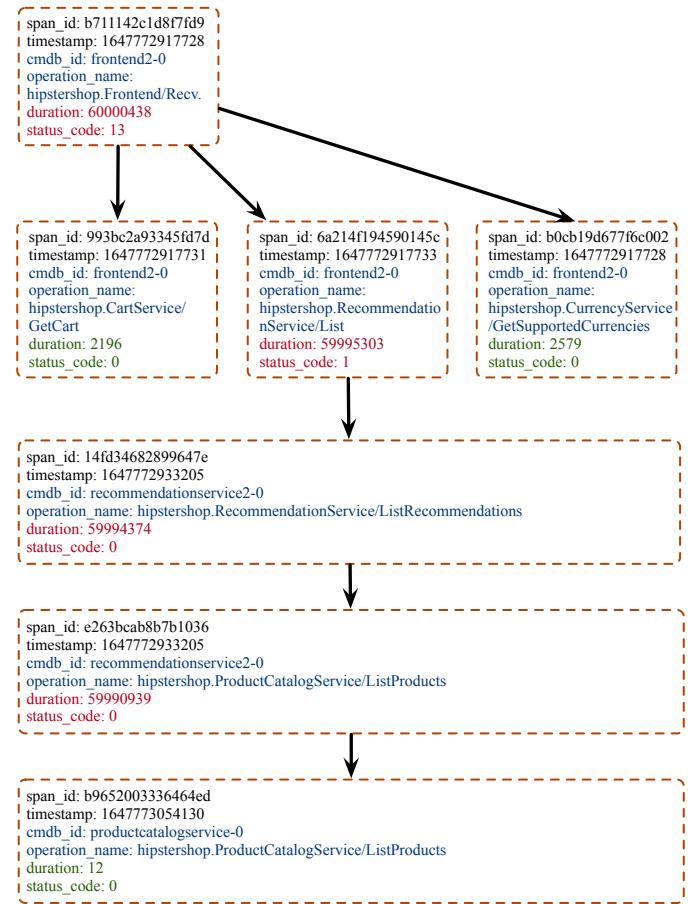


Fig. 2. Trace Graph Example-1

For instance, we extracted a typical case from the real-world AIOPS 2022 dataset, as shown in Figure 2. In this case, a request with the cmdb\_id 'frontend2-0' experienced severe timeouts, ultimately resulting in an error (status\_code = 13). As SREs, we first analyzed the trace data for this request and identified three primary operations: CartService/GetCart, RecommendationService/List, and CurrencyService/GetSupportedCurrencies. Notably, only the RecommendationService operation exhibited a timeout, indicating that it was likely responsible for the overall request failure.

We then conducted a deeper investigation into the downstream services and found that the RecommendationService operation with cmdb\_id 'recommendationservice2-0' also experienced a timeout. Further analysis revealed that the subsequent downstream operation, ProductCatalogService, under the same cmdb\_id, also timed out. However, a later operation on the component with cmdb\_id 'productcatalogservice-0' did not exhibit any issues.

This recursive investigation process effectively narrows down the potential root causes to the components with cmdb\_id 'frontend2-0' and 'recommendationservice2-0', significantly reducing the search space for identifying the underlying performance issue.

**Summary.** By analyzing trace data, SREs narrow down the search space for the root cause by starting from the entry span of a high-duration request and recursively examining downstream operations until the anomalous component is identified.

### B. Multi-Dimensional Expansion

The previous analysis narrowed the search space by focusing on cmdb\_id, which represents individual pods. However, identifying problematic cmdb\_ids alone is insufficient for pinpointing all root causes. Therefore, SREs expand their analysis beyond individual pods by considering operation\_name, which helps identify the associated services. From the recursive analysis above, it is evident that both recommendationservice and productcatalogservice could also be potential root causes.

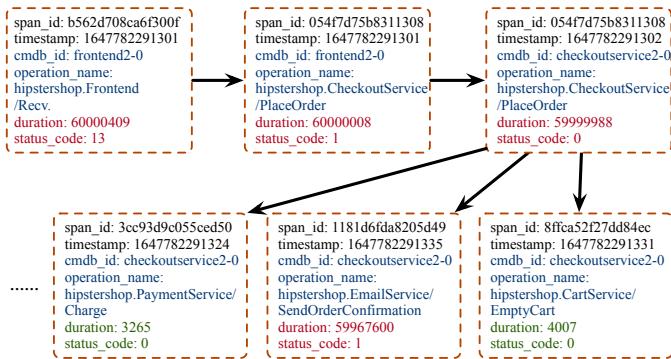


Fig. 3. Trace Graph Example-2

A more evident example is shown in Figure 3. In this trace, the root cause is clearly linked to the EmailService operation under cmdb\_id 'checkoutservice2-0'. This suggests that further investigation into emailservice is necessary for a deeper analysis.

Beyond service-level expansion, SREs also consider the underlying physical infrastructure. For instance, in both traces above, 'recommendationservice2-0' and 'checkoutservice2-0' run on node-5, which can be readily obtained in real-time from metrics data. This indicates that further investigation into the physical machine node-5 is required to assess its potential contribution to the issue.

**Summary.** SREs systematically expand the scope of root cause analysis across multiple dimensions—correlating anomalous cmdb\_ids with associated services and underlying physical infrastructure—to comprehensively identify all potential sources of issues.

### C. Cross-Modal Reasoning

Through the analysis of recursiveness and multi-dimensional expansion, SREs have identified potential root causes across three dimensions: pod, service, and node. However, the exact root cause remains uncertain. At this stage, SREs often integrate other modalities of data, with metrics data being the most commonly used, to pinpoint the final root cause.

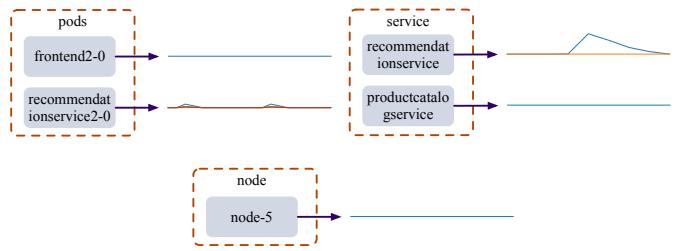


Fig. 4. Metrics Inspection of Trace Graph Example-1

For example, in Figure 2, the potentially problematic pod is 'frontend2-0', and 'recommendationservice2-0', the potentially problematic services are 'recommendationservice' and 'productcatalogservice', and the potentially problematic node is 'node-5'. By analyzing the metrics data, as shown in Figure 4, it becomes clear that there were no fluctuations in 'frontend2-0', 'productcatalogservice', or 'node-5'. However, both 'recommendationservice2-0' and 'recommendationservice' exhibited fluctuations, with 'recommendationservice' showing more significant variations. The fluctuations in 'recommendationservice2-0' were propagated from 'recommendationservice'. Ultimately, the root cause is identified as 'recommendationservice', which aligns with the ground truth—network resource packet corruption in the Kubernetes container.

**Summary.** SREs differentiate the actual root cause from other potential anomalies by analyzing other modalities of data, particularly metrics data, and identifying the component with the most significant or earliest fluctuation as the true root cause.

## IV. RCLAGENT

Our empirical study demonstrates the process by which human SREs localize the root cause of abnormal requests. However, the time and effort of human experts are valuable, and large language models can effectively simulate this process for automated root cause localization.

Therefore, in this section, we introduce **RCLAgent**, an adaptive method for root cause localization for microservice systems through multi-agent recursion-of-thought. Figure 5 illustrates the architecture of RCLAgent.

RCLAgent consists of two types of agents: data agents and thought agents. Data agents are responsible for data retrieval and processing (currently including the Trace Agent, Metric Agent, and Format Agent), while thought agents focus on in-depth reasoning based on the provided data (including the Recursion Agent and Cross-Modal Agent). The coordination between these agents is managed by a central coordinator, which operates through three key phases: initial reasoning, critical reflection, and final review.

### A. Data Agents

Data agents provide various types of data to other agents based on input, along with preprocessing and filtering capabilities. They can encompass any form of data retrieval and

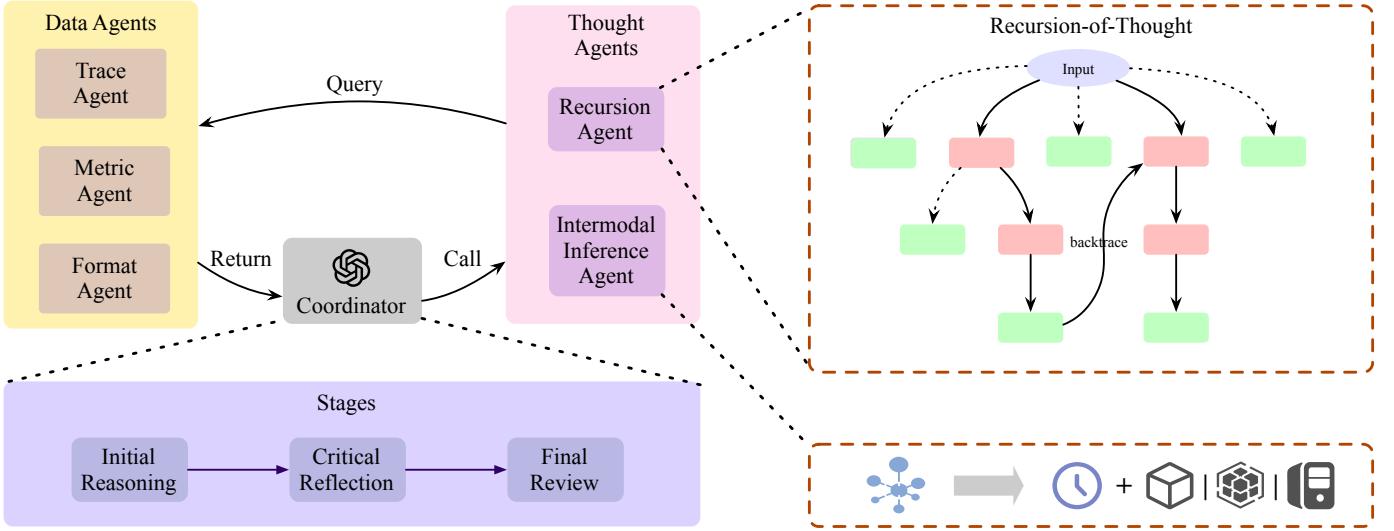


Fig. 5. Architecture of RCLAgent

processing functions. In this work, we primarily utilize three agents: the Trace Agent, Metrics Agent, and Format Agent. However, in practical applications, the number and types of agents can be flexibly adjusted as needed.

1) *Trace Agent*: Trace data is the most fundamental source for root cause localization, as it records the sequence of calls between various services. However, since each request generates a complete invocation path, the volume of trace data can be enormous, making it difficult for large language models to process such extensive context. To address this, the trace agent is designed to filter and retrieve only the relevant subset of trace data.

$$T(s) = \{\langle t, s', svc, op, d, \sigma \rangle \mid s' \in \mathcal{C}(s)\} \quad (2)$$

Formally, given a span identifier  $s$ , the trace agent returns a set of child spans along with their associated metadata. We denote this function as Equation 2, where  $t$  denotes the timestamp,  $s'$  represents the child span identifier,  $svc$  is the service name,  $op$  is the operation name,  $d$  is the duration,  $\sigma$  is the status code, and  $\mathcal{C}(s)$  is the set of all child spans for the given span  $s$ .

2) *Metric Agent*: Metrics data represent the runtime state of system components. In a mature microservices environment, there are thousands of distinct metrics continuously recording values, often resulting in a data volume that exceeds that of trace data. Moreover, our empirical study reveals that during system anomalies, most metrics remain stable without significant fluctuations. Consequently, we have developed a specialized Metric Agent that selectively retrieves only those metrics exhibiting notable deviations within a predefined time window for LLM analysis.

$$|m(t) - \mu_m| > n \times \sigma_m \quad (3)$$

Formally, given an input timestamp  $t_0$  and a target key component  $C$  (e.g., a pod or service), let  $\mathcal{M}(C)$  denote the set of metrics associated with  $C$  and its related components (including the node corresponding to a pod). For each metric

$m \in \mathcal{M}(C)$ , let  $\mu_m$  and  $\sigma_m$  represent its historical mean and standard deviation, respectively. The Metric Agent examines the metric values over the interval  $[t_0 - \delta, t_0 + \delta]$  and applies an  $n$ -sigma test: if there exists a time  $t$  in this interval, as defined in Equation 3, an anomaly is detected. In such cases, the agent returns the corresponding fluctuation data over a period of length  $delta$ , as formulated in Equation 4.

$$Q(t_0, \delta, C) = \left\{ m(t) \mid \begin{array}{l} m \in \mathcal{M}(C), t \in [t_0 - \delta, t_0 + \delta], \\ |m(t) - \mu_m| > n \times \sigma_m \end{array} \right\} \quad (4)$$

3) *Format Agent*: Large language models generate extensive natural language outputs that often contain key information. However, this abundance of unstructured text can hinder subsequent automated processing, making it difficult to produce a definitive final result. To address this issue, we designed a Format Agent specifically to summarize the generated text and output it in a standardized format. For instance, when presenting the final results, the Format Agent ensures that exactly two fields are output: one for the root cause (`root_cause`) and one for the corresponding explanation (`reason`).

## B. Thought Agents

Thought agents are designed based on insights from our empirical study, specifically leveraging recursiveness, multi-dimensional expansion, and cross-modal reasoning. Since large language models inherently possess the ability to extract key information from vast amounts of data without excessive deliberation, this section focuses primarily on the design of recursiveness and cross-modal reasoning.

1) *Intermodal Inference Agent*: Trace data and metrics data exhibit a semantic gap, making direct correlation challenging. To address this, the *Intermodal Inference Agent* is designed to bridge the gap between trace analysis and metrics querying by synthesizing query parameters from multi-modal contextual data. This agent extracts the most relevant keywords (potential

root causes) and temporal windows from the trace context. These parameters are then utilized to query the metrics data, ensuring that only those metrics corroborating the anomalies detected in trace data are retrieved.

$$q = I(X) = (K, \tau) \quad (5)$$

Formally, let  $X$  denote the contextual embedding derived from trace analysis. The intermodal inference agent computes a query tuple  $q = (K, \tau)$ , as shown in Equation 5, where  $K$  represents the set of key metric identifiers (i.e., keywords), and  $\tau$  denotes the inferred time window during which anomalies are expected. The function  $I(\cdot)$  encapsulates the agent's inference process, modeling the complex interactions between different data modalities.

2) *Recursion Agent*: As previously discussed, professional SREs employ a recursive approach when using trace data for root cause localization. To replicate this process, we designed a dedicated Recursion Agent that follows a recursion-of-thought paradigm.

Unlike traditional chain-of-thought reasoning, which follows a linear progression, recursion-of-thought relies on generating step-by-step reasoning instructions based on the information from the previous step. Each generated instruction triggers subordinate agents to retrieve the necessary data for further analysis. Moreover, if a particular analytical path yields no further potential for identifying a root cause, the agent automatically backtraces to a previously encountered candidate that has not been thoroughly inspected, and continues the recursive analysis.

In this paper, the recursion-of-thought mechanism is primarily applied to trace data analysis, with metrics data verification used to confirm findings. This process involves the coordinated interaction of all agents in our framework, as illustrated in Algorithm 1.

Specifically, the process begins with the trace agent identifying the entry span of a high-duration request, which serves as the starting point for recursive analysis. At each step, an LLM-based reasoning instruction is generated based on the current candidate span. The trace agent then retrieves the associated trace data, and if the candidate is deemed a potential root cause, the intermodal inference agent synthesizes query parameters (keywords and a temporal window) from the trace context. These parameters are subsequently passed to the metrics agent, which retrieves the corresponding metrics data to verify whether the observed anomaly is corroborated. If the metrics analysis confirms the anomaly, the candidate is added to the potential root cause set; otherwise, the algorithm expands the search by exploring the candidate's child spans. Moreover, if a particular analysis path fails to yield a viable candidate, the system automatically backtraces to previously uninspected candidates for further investigation. Finally, the format agent consolidates and standardizes the identified potential root causes into a structured final output.

### C. Coordinator

Although the recursion agent has outlined the invocation relationships between agents from an algorithmic perspective,

---

### Algorithm 1 Recursion-of-Thought Algorithm

---

**Require:** Trace data  $T$  for a high-duration request.

**Ensure:** Formatted potential root cause(s)  $R_f$ .

```

1: Initialize candidate set  $Q \leftarrow \{T_{\text{entry}}\}$ 
2: Initialize potential root cause set  $R \leftarrow \emptyset$ .
3: while  $Q \neq \emptyset$  do
4:   Remove a candidate span  $s$  from  $Q$ .
5:   Generate reasoning instruction  $I \leftarrow f(s)$ 
6:   Retrieve associated data  $D \leftarrow \text{TraceAgent}(I)$ .
7:   if  $s$  is deemed a potential root cause based on  $D$  then
8:     Compute query tuple  $q \leftarrow I_{\text{intermodal}}(X) = (K, \tau)$ 
9:     Retrieve metrics data  $M \leftarrow A_{\text{metric}}(q)$ .
10:    if Metrics analysis confirms anomaly for  $s$  then
11:       $R \leftarrow R \cup \{s\}$ 
12:    else
13:      Discard  $s$ 
14:    end if
15:  else
16:    Let  $C$  be the set of child spans of  $s$ .
17:     $Q \leftarrow Q \cup C$ .
18:  end if
19:  if  $s$  yields no further viable path and backtracing is
feasible then
20:    Backtrace to a previously uninspected candidate in
 $Q$ .
21:  end if
22: end while
23:  $R_f \leftarrow \text{FormatAgent}(R)$ 
24: return  $R_f$ .

```

---

the actual process of querying data agents, processing their returned results, and transmitting the information to thought agents for further execution is orchestrated by the coordinator. Moreover, beyond managing these interactions, the coordinator also oversees the execution of Algorithm 1 across three distinct phases, ensuring iterative coordination throughout the process.

1) *Initial Reasoning*: This stage is responsible for initializing the entire inference process. Given an initial task, the model is induced to consider the problem from a recursive perspective and begin a preliminary assessment of potential root causes. To expedite reasoning at this stage, we deliberately hide the presence of the Metrics Agent.

#### System:

You are a software operations engineer. Your task is to systematically diagnose and identify the root cause of software failures.

#### User:

Please read the following root trace and identify the corresponding root cause service.

A possible approach is to recursively search for the traces you suspect have issues.  $\{T_{\text{entry}}\}$

You have the following agents to call:  $\{A \setminus \{A_{\text{metric}}\}\}$

Fig. 6. The Prompt for Initial Reasoning

As shown in Figure 6, the initial prompt, denoted as  $P_0$ , is constructed from the entry trace  $T_{\text{entry}}$  and a set of available agents  $A \setminus \{A_{\text{metric}}\}$ . The initial reasoning state  $S_0$  is formally defined as Equation 6, where  $\Phi(\cdot)$  represents the inference process executed by the large language model.

$$S_0 = \Phi(P_0(T_{\text{entry}}, A \setminus \{A_{\text{metric}}\})) \quad (6)$$

2) *Critical Reflection:* Initial reasoning can quickly yield a rough judgment of the root cause; however, our experiments reveal that the model tends to terminate its analysis prematurely. For example, as shown in Figure 3, during the initial reasoning phase the model often identifies the problematic pod as `checkoutservice2-0` after just two recursive trace queries. In reality, if the model continues the recursion for a few more steps, it discovers that the actual anomaly originates from `emailservice`. This observation highlights the need for a deeper, more rigorous analysis to avoid early termination and misidentification of the true root cause.

### Step 1:

We think it's better to inspect more deeper into the trace tree. So, even you have already define the root cause, you must use the following tool to inspect more deeper to confirm that there are no deeper root cause.

You have the following agents to call:  $\{A \setminus \{A_{\text{format}}\}\}$

### Step 2:

Please continue to further identify the root cause service. You may inspect deeper by Trace Agent or combine with Metrics Agent to confirm the root cause.

If you use the Trace Agent, we often find that the root cause originates from a specific downstream trace. If you have already define the root cause service, just call the Format Agent.

You have the following agents to call:  $\{A\}$

Fig. 7. The Prompt for Critical Reflection

To address this shortcoming, we introduce a critical reflection stage that forces the model to conduct further investigation before finalizing its decision. In this stage, the model is explicitly prompted to delve deeper into the trace tree, even if an initial root cause candidate has been identified. As depicted in Figure 7, the model receives a prompt that instructs it to either continue its investigation using the Trace Agent or to incorporate insights from the Metrics Agent for additional verification. This additional layer of scrutiny helps ensure that the model does not overlook deeper anomalies that may be hidden beneath the initial candidate.

$$S_1 = \mathcal{R}(R_0) = \bigcup_{i=1}^n \mathcal{Q}(Q_{i-1}, A) \quad (7)$$

Formally, let  $R_0$  denote the initial set of potential root causes identified during the initial reasoning stage. The critical reflection stage, denoted as  $S_1$ , refines  $R_0$  by recursively expanding the candidate set. Specifically, we define the refined set  $R_f$  as Equation 7, where  $\mathcal{Q}(Q_{i-1}, A)$  represents the

expansion of the candidate set  $Q_i$  via agent  $A$  (such as the Trace or Metrics Agent), and  $\mathcal{R}(\cdot)$  encapsulates the recursive refinement process.

3) *Final Review:* After the critical reflection stage, the RCLAgent performs in-depth inspections of both trace and metrics data. However, in practice, large language models tend to focus more on the recent context when generating results. As a result, potential root causes that appear earlier in the analysis might be overlooked. For example, as shown in Figure 2, after deeper inspection, the potential root cause could be identified as `productcatalogservice`. However, the actual root cause in this case lies in the shallower level, `recommendationservice`, which might be missed if the model prematurely concludes the investigation. This scenario illustrates that, while the initial reasoning might suggest `recommendationservice`, the critical reflection stage could indicate `productcatalogservice`, leading to conflicting conclusions.

To address this issue, we introduced a final review stage, where only the Format Agent is invoked. This agent allows the model to review the entire reasoning process, starting from the initial stage, and consolidate its findings to determine the final root cause. By revisiting the prior reasoning, the model ensures that any earlier potential root causes are not overlooked, and the final decision is based on the entire diagnostic process. This structured review helps the model arrive at the most accurate and reliable root cause conclusion.

### Context:

Please rethink the above think process and return the correct root cause and reason by the  $A_{\text{format}}$  agent.

### Input:

{Think Process}

Fig. 8. The Prompt for Final Review

As shown in Figure 8, the final review stage is implemented as a consolidation process where the reasoning results from both the initial reasoning and critical reflection stages are combined. This is mathematically represented by the Equation 8, where  $\mathcal{R}(R_0)$  represents the initial root cause set from the initial reasoning, and  $\mathcal{R}(S_1)$  is the refined set from the critical reflection stage.

$$R_f = A_{\text{format}} (\mathcal{R}(R_0) \cup \mathcal{R}(S_1)), \quad (8)$$

## V. EVALUATION

To evaluate RCLAgent, we conduct a series of experimental studies to investigate the following research questions:

- **RQ1:** How accurate is RCLAgent in root cause localization when compared with baseline approaches?
- **RQ2:** How does the choice of different LLM backbones affect the root cause localization performance of RCLAgent?
- **RQ3:** What is the contribution of each stage of RCLAgent to its overall accuracy?

TABLE I  
OVERALL EVALUATION RESULTS COMPARED WITH NON-LLM-BASED METHODS (RECALL@K)

Dataset	CRISP			TraceContrast			TraceRank			MicroRank			RUN			MicroScope			RCLAgent
	R1	R5	R10	R1	R5	R10	R1	R5	R10	R1	R5	R10	R1	R5	R10	R1	R5	R10	-
A	0.00	28.14	28.14	1.69	22.71	24.07	0.68	0.68	15.93	0.00	26.10	34.93	3.73	13.22	42.03	9.68	36.92	68.46	<b>71.13</b>
B	0.00	60.54	61.88	56.05	75.78	75.78	75.78	75.78	75.78	0.90	60.54	76.23	0.00	0.00	0.00	1.47	7.35	14.22	<b>78.57</b>
$\Gamma$	2.51	48.62	51.62	46.35	78.92	88.98	23.95	43.35	56.65	15.45	80.84	89.70	23.83	24.55	24.55	20.95	59.76	81.18	<b>90.24</b>
$\Delta$	1.15	52.87	56.05	2.19	2.92	3.13	59.44	60.06	63.48	2.82	3.34	3.34	0.42	14.18	17.52	7.61	20.75	26.07	<b>64.34</b>
E	15.20	62.33	63.06	33.26	34.14	35.68	23.79	51.76	59.69	27.09	35.68	35.68	0.00	6.39	31.28	8.81	37.44	54.19	<b>65.42</b>
Z	4.68	40.76	45.32	4.81	11.65	17.22	34.94	35.70	39.75	4.43	16.96	17.47	0.13	12.91	39.49	6.86	37.98	63.39	<b>72.41</b>

### A. Experimental Setup

1) *Dataset*: Our experiments are conducted using the AIOPS 2022 companion dataset, which is a large-scale, real-world dataset collected from an actively running, mature microservices-based e-commerce system. This dataset is divided into six distinct subsets. For the sake of clarity and ease of reference in this paper, we label these six subsets as **A**, **B**,  $\Gamma$ ,  $\Delta$ , **E**, and **Z**.

2) *Baseline Approaches*: We compared TraceContrast with the following eight root cause localization approaches, which can be categorized into three groups: trace-based, metrics-based, and LLM-based methods.

**Trace-based**: CRISP [24] represents traces as critical paths and applies a lightweight heuristic to identify root cause instances. TraceContrast [25] utilizes sequence representations, contrastive sequential pattern mining, and spectrum analysis to localize multi-dimensional root causes. TraceRank [23] combines spectrum analysis with a PageRank-based random walk algorithm to pinpoint anomalous services. MicroRank [22] constructs a trace coverage tree to capture dependencies between requests and service instances, leveraging the PageRank algorithm to score potential root causes.

**Metrics-based**: RUN [21] employs time series forecasting for neural Granger causal discovery and integrates a personalized PageRank algorithm to efficiently recommend the top-k root causes. Microscope [19] builds causality graphs and utilizes a depth-first search strategy to detect front-end anomalies.

**LLM-based**: mABC [26] proposes a multi-agent, blockchain-inspired collaboration framework where multiple LLM-based agents follow a structured workflow and collaborate through blockchain-inspired voting mechanisms. Additionally, we implement a CoT-based Approach [31] which implements a naive chain-of-thought (CoT) reasoning method for root cause localization.

3) *Evaluation Metrics*: We use the top-k recall (Recall@k) and mean reciprocal rank (MRR) to evaluate the accuracy of root cause localization following existing works [2], [19], [22], [25].

- **Recall@k**: Measures the likelihood that the true root cause appears within the top-k results in the ranked list. Specifically, it indicates whether the root cause is found within the first  $k$  predictions. In this paper, we evaluate Recall@1, Recall@5, and Recall@10.

- **MRR**: is the multiplicative inverse of the rank of the root cause in the result list. If the root cause is not included in

the top-10 result list, the rank can be regarded as positive infinity. Given a set of fault instances  $A$ ,  $Rank_i$  is the  $i$  rank of the root cause in the returned list of the  $i$ th fault instance, MRR is calculated by Equation 9.

$$MRR = \frac{1}{|A|} \sum_{i=1}^{|A|} \frac{1}{Rank_i} \quad (9)$$

Since many existing methods are limited to localizing the root cause only at the pod level, we assume that if these methods correctly identify the pod, the corresponding service to which the pod belongs is considered the root cause, and we treat such predictions as correct.

4) *Implementation and Settings*: We implement RCLAgent in Python 3.10. Unless otherwise specified, we use Claude-3.5 Sonnet as our LLM engine and set the n-sigma threshold in the metrics agent to  $n = 3$ . All experiments are conducted on a Linux server equipped with 24 Intel(R) Xeon(R) processors (2.90GHz), 400GB RAM, and two A800 GPUs with 80GB of GPU memory, running CentOS 8.

### B. Overall Evaluation

We first compare RCLAgent with trace-based methods and metrics-based methods. As shown in Table I, all of these SOTA methods rely on initial service relationships and predefined causal graph construction approaches. Therefore, they require collecting multiple requests and constructing a causal graph before ranking potential root causes based on their weights. In contrast, RCLAgent only analyzes a single request and directly identifies the exact root cause. Thus, the results presented for RCLAgent in the table correspond to Recall@1 (R1), since it determines the exact root cause with just one request.

From the table, we observe that even when analyzing only a single request to identify the exact root cause, RCLAgent outperforms the SOTA methods' Recall@10 (R10) results by approximately 2% to 9%. Furthermore, compared to other methods achieving the same Recall@1 (R1) performance, RCLAgent shows an average improvement of 30.44%. This demonstrates the superior effectiveness of RCLAgent, even when working with a single request.

Next, we compare the performance of RCLAgent with LLM-based methods. As shown in Table II, when evaluating on Recall@1 (R1), RCLAgent outperforms mABC by an average of 15.60% in accuracy. Additionally, mABC combines log data analysis alongside trace and metric data, and with a higher number of agents involved, its inference speed is slower

TABLE II  
EVALUATION RESULTS COMPARED WITH LLM-BASED METHODS

Approach	A	B	$\Gamma$	$\Delta$	E	Z
CoT	11.90	21.43	39.95	16.25	20.20	25.47
mABC	51.87	62.22	74.23	56.35	52.08	51.65
RCLAgent	<b>71.13</b>	<b>78.57</b>	<b>90.24</b>	<b>64.24</b>	<b>65.42</b>	<b>72.41</b>

compared to RCLAgent. In comparison with the naive CoT strategy, RCLAgent demonstrates a much more significant advantage, outperforming it by up to 60% on certain datasets. This result highlights that root cause localization is not a problem that can be easily solved by merely using LLMs, further emphasizing the necessity of the design of RCLAgent.

### C. Group Ranking Evaluation

The previous experiments demonstrated that RCLAgent, by analyzing a single request to provide the exact root cause, can outperform SOTA methods. To further assess the superiority of RCLAgent, we introduced a naive majority voting algorithm. Specifically, for a set of multiple requests within a time window, RCLAgent analyzes each request to identify its corresponding root cause. The root causes are then aggregated using a majority voting mechanism to determine the final root cause, in a manner analogous to SOTA methods, which rely on constructing a causal graph and ranking potential root causes.

Formally, the majority voting procedure can be described as follows. Let  $\{r_1, r_2, \dots, r_k\}$  represent the root causes identified for  $k$  requests within a time window. The majority vote is given by Equation 10, where  $\hat{R}$  is the final predicted root cause,  $\mathbb{I}(r_i = r)$  is the indicator function that equals 1 if the  $i$ -th root cause  $r_i$  matches the candidate root cause  $r$ , and 0 otherwise. The root cause  $\hat{R}$  is the one that receives the highest number of votes.

$$\hat{R} = \operatorname{argmax}_r \left( \sum_{i=1}^k \mathbb{I}(r_i = r) \right) \quad (10)$$

TABLE III  
MRR COMPARISON

Approach	A	B	$\Gamma$	$\Delta$	E	Z
CRISP	8.27	20.13	18.13	17.34	31.08	17.14
TraceConstruct	13.07	<b>65.74</b>	<b>58.55</b>	2.48	33.77	8.15
TraceRank	6.26	<b>76.76</b>	34.41	<b>61.54</b>	<b>35.79</b>	<b>38.36</b>
MicroRank	11.38	18.12	38.10	2.98	30.81	9.15
RUN	11.72	3.12	25.65	5.62	7.58	8.95
MicroScope	<b>23.76</b>	4.55	37.46	13.24	21.38	21.33
RCLAgent	<b>77.63</b>	<b>78.57</b>	<b>94.44</b>	<b>84.98</b>	<b>76.02</b>	<b>78.31</b>

After applying the aforementioned majority voting algorithm, we conducted a detailed analysis of RCLAgent and the SOTA methods in terms of MRR, as shown in Table III. From the table, it is evident that RCLAgent outperforms all the SOTA methods in terms of MRR. On average, RCLAgent achieves a 32.53% improvement over the second-best method. This demonstrates the superiority of RCLAgent in the group ranking evaluation. Its excellent performance in

Recall@1 (R1) ensures that even with a simple majority voting algorithm, RCLAgent delivers impressive results.

### D. LLM Backbone Impact

Next, we address the second research question. To evaluate the performance of RCLAgent across different LLM backbones, we conducted experiments with four additional state-of-the-art LLMs beyond Claude-3.5-Sonnet: DeepSeek-R1-Qwen, which is the Qwen2.5-32B model fine-tuned on distilled data released by DeepSeek; Qwen-2.5-Max and Qwen-2.5-Plus, both of which are closed-source models accessed via API; and Llama-3.1-70B, the original open-source version released by Meta.

TABLE IV  
RCLAGENT WITH DIFFERENT LLMs

Model	A	B	$\Gamma$	$\Delta$	E	Z
Claude-3.5-sonnet	<b>71.13</b>	<b>78.57</b>	<b>90.24</b>	<b>64.34</b>	<b>65.42</b>	<b>72.41</b>
DeepSeek-R1-qwen	33.33	58.73	67.55	69.17	45.06	61.36
Qwen-2.5-max	13.17	21.23	33.66	13.17	33.66	13.68
Qwen-2.5-plus	<b>39.79</b>	<b>70.45</b>	52.78	49.48	35.46	52.06
Llama-3.1-70B	7.92	22.42	17.51	18.14	6.97	8.76

As shown in Table IV, the performance of RCLAgent is closely tied to the reasoning capabilities of the underlying LLM backbone. In our experiments, Claude-3.5-Sonnet consistently achieved the best results, followed by DeepSeek-R1-Qwen. However, due to its relatively smaller size (32B), DeepSeek-R1-Qwen underperformed on datasets A and B compared to Qwen-2.5-Plus. Overall, RCLAgent with Claude-3.5-Sonnet outperformed the second-best model by an average of 14.79%.

Additionally, we experimented with other models such as GPT-4o and Llama-3-70B, but their performance was even lower than the weakest models listed in Table IV. These findings highlight that RCLAgent effectively leverages the semantic understanding and logical reasoning capabilities of large language models, achieving better results when paired with more powerful inference engines.

### E. Ablation Study

Then, to address the third research question, we recorded the complete reasoning path of RCLAgent during the experiments. Additionally, we extracted intermediate results at three distinct stages and evaluated their effectiveness in root cause localization.

As shown in Figure 9, CR denotes Critical Reflection, and FR represents Final Review. The figure demonstrates that these three stages significantly enhance RCLAgent's root cause localization performance. For instance, on dataset A, applying Critical Reflection improved the results of Initial Reasoning by approximately 34.51%. Similarly, Final Review consistently contributed to performance gains, albeit to a lesser extent than Critical Reflection. On datasets B and  $\Gamma$ , Final Review further improved the results by approximately 6% beyond the gains from Critical Reflection.

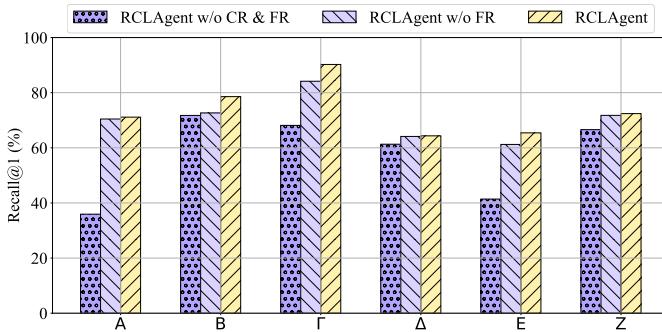


Fig. 9. Ablation Experiment

Overall, Initial Reasoning alone achieved an R1 (Recall@1) score of approximately 57.52%. Incorporating Critical Reflection led to an average improvement of 13.31%, while Final Review provided an additional 2.96% boost. These findings highlight the effectiveness and rationality of RCLAgent's three-stage design.

## VI. THREATS TO VALIDITY

Despite the promising results demonstrated by RCLAgent, several limitations should be acknowledged. First, the implementation and configuration of baseline approaches: CRISP, MicroRank, and mABC have publicly available source code, which we directly utilized. For other state-of-the-art methods, we implemented them based on the descriptions in their respective papers. After implementation, we fine-tuned the results and selected the best configurations for all baselines through experimentation. Second, our evaluation primarily relies on the AIOps 2022 dataset, which, while extensive and derived from real-world microservice systems, may not fully capture the diversity of failure scenarios encountered in different operational contexts. Third, although our empirical study benefits from insights provided by multiple professional SREs, the inherent subjectivity in their diagnostic processes may introduce bias.

## VII. RELATED WORK

### A. Trace-Based Root Cause Localization

With the advancement of distributed tracing and its supporting infrastructure, researchers have explored various trace-based techniques for root cause localization. Several studies [1], [32], [33] have analyzed the role of distributed tracing in diagnosing failures in large-scale microservices systems, highlighting the importance of automated trace analysis for effective root cause localization.

Recently, machine learning techniques have been widely adopted for trace-based root cause localization. For instance, Zhou et al. [34] proposed MEPFL, a supervised learning-based approach. Gan et al. [35] introduced Seer, a model leveraging CNNs and LSTMs for root cause analysis. Liu et al. [36] proposed TraceAnomaly, an unsupervised approach for anomaly detection in traces. Additionally, Gan et al. [37] developed Sage, which utilizes graph neural networks for root cause localization.

Beyond machine learning-based methods, some researchers have extended spectrum analysis techniques to improve the practicality of trace-based root cause localization. For example, Yu et al. [22] proposed MicroRank, which was later extended into TraceRank [23] by integrating spectrum analysis with a random walk-based method. Li et al. [38] combined spectrum analysis with frequent pattern mining to identify root cause service instances. More recently, Zhang et al. [25] introduced TraceConstruct, which leverages sequence representations along with contrast sequential pattern mining and spectrum analysis to efficiently localize multi-dimensional root causes.

### B. LLM-based Failure Management

Large language models, with their advanced semantic understanding and logical reasoning capabilities, have significantly improved the field of failure management [2] and are increasingly becoming a focal point of research. Numerous LLM-based approaches have been proposed to address various aspects of failure management, including anomaly detection, failure diagnosis, and automated mitigation [39]–[65].

Some studies have developed foundation models specifically for failure management. For example, Lag-Llama [39], Timer [40], and TimesFM [41] pretrain foundation models for metrics-based anomaly detection. Similarly, ShellGPT [42] trains a model capable of automatically generating shell scripts for automated mitigation.

Other approaches adopt fine-tuning strategies to tailor LLMs for failure management tasks. For instance, AnomalyLLM [43] and UniTime [44] employ full fine-tuning for anomaly detection, while OWL [45] and LogLM [46] leverage parameter-efficient fine-tuning techniques to build log analysis models.

Since these fine-tuning approaches require significant computational resources and time, an increasing number of methods, including our work, rely on prompt-based techniques. For example, RCACopilot [47] and Xpert [48] utilize in-context learning (ICL) to structure diagnostic processes, ensuring accurate root cause analysis. LM-PACE [49] applies chain-of-thought (CoT) reasoning to enhance GPT-4's ability to analyze incident reports, while Hamadanian et al. [50] extend this approach to generate mitigation solutions from incident reports. Additionally, RAGLog [51] and LogRAG [52] use retrieval-augmented generation (RAG) to enhance log-based anomaly detection through historical log retrieval.

## VIII. CONCLUSION

In this paper, we study the problem of adaptive root cause localization for microservice systems. Initially, we conduct a comprehensive study on how SREs localize the root causes of an abnormal request. This study reveals that human root cause analysis exhibits three key characteristics: recursive-ness, multi-dimensional expansion, and cross-modal reasoning. Based on our study, we introduce RCLAgent, an adaptive root cause localization method for microservice systems that leverages a multi-agent recursion-of-thought framework. RCLAgent employs a novel recursion-of-thought approach to guide the LLM's reasoning process, effectively utilizing data

from multiple agents and tool-assisted analysis to achieve accurate root cause localization. Our experimental evaluations on various public datasets demonstrate that RCLAgent achieves superior performance by analyzing a single request, surpassing state-of-the-art methods that rely on multiple requests for analysis.

In future work, we will further explore how to achieve more accurate and efficient root cause localization using smaller-scale models. Additionally, we are considering extending our approach to cover the entire failure management process.

## REFERENCES

- [1] X. Zhou, X. Peng, T. Xie, J. Sun, C. Ji, W. Li, and D. Ding, “Fault analysis and debugging of microservice systems: Industrial survey, benchmark system, and empirical study,” *IEEE Transactions on Software Engineering*, vol. 47, no. 2, pp. 243–260, 2018.
- [2] L. Zhang, T. Jia, M. Jia, Y. Wu, A. Liu, Y. Yang, Z. Wu, X. Hu, P. S. Yu, and Y. Li, “A survey of aiops for failure management in the era of large language models,” *arXiv preprint arXiv:2406.11213*, 2024.
- [3] L. Zhang, T. Jia, M. Jia, Y. Wu, A. Liu, Y. Yang, Z. Wu, X. Hu, P. Yu, and Y. Li, “A survey of aiops in the era of large language models,” *ACM Computing Surveys*, 2025.
- [4] N. C. Mendonça, P. Jamshidi, D. Garlan, and C. Pahl, “Developing self-adaptive microservice systems: Challenges and directions,” *IEEE Software*, vol. 38, no. 2, pp. 70–79, 2019.
- [5] M. Waseem, P. Liang, M. Shahin, A. Di Salle, and G. Márquez, “Design, monitoring, and testing of microservices systems: The practitioners’ perspective,” *Journal of Systems and Software*, vol. 182, p. 111061, 2021.
- [6] L. Zhang, T. Jia, M. Jia, H. Liu, Y. Yang, Z. Wu, and Y. Li, “Towards close-to-zero runtime collection overhead: Raft-based anomaly diagnosis on system faults for distributed storage system,” *IEEE Transactions on Services Computing*, 2024.
- [7] L.-Z. Zhang, X.-D. Huang, Y.-K. Wang, J.-L. Qiao, S.-X. Song, and J.-M. Wang, “Time-tired compaction: An elastic compaction scheme for lsm-tree based time-series database,” *Advanced Engineering Informatics*, vol. 59, p. 102224, 2024.
- [8] Y. Kang, X. Huang, S. Song, L. Zhang, J. Qiao, C. Wang, J. Wang, and J. Feinauer, “Separation or not: On handing out-of-order time-series data in leveled lsm-tree,” in *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 2022, pp. 3340–3352.
- [9] L. Zhang, T. Jia, M. Jia, Y. Li, Y. Yang, and Z. Wu, “Multivariate log-based anomaly detection for distributed database,” in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 4256–4267.
- [10] L. Zhang, T. Jia, K. Wang, M. Jia, Y. Yang, and Y. Li, “Reducing events to augment log-based anomaly detection models: An empirical study,” in *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 2024, pp. 538–548.
- [11] D. Wang, Z. Chen, J. Ni, L. Tong, Z. Wang, Y. Fu, and H. Chen, “Inter-dependent causal networks for root cause localization,” in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 5051–5060.
- [12] S. Zhang, S. Xia, W. Fan, B. Shi, X. Xiong, Z. Zhong, M. Ma, Y. Sun, and D. Pei, “Failure diagnosis in microservice systems: A comprehensive survey and analysis,” *ACM Transactions on Software Engineering and Methodology*, 2024.
- [13] Q. Yu, N. Zhao, M. Li, Z. Li, H. Wang, W. Zhang, K. Sui, and D. Pei, “A survey on intelligent management of alerts and incidents in it services,” *Journal of Network and Computer Applications*, p. 103842, 2024.
- [14] Y. Sun, Z. Lin, B. Shi, S. Zhang, S. Ma, P. Jin, Z. Zhong, L. Pan, Y. Guo, and D. Pei, “Interpretable failure localization for microservice systems based on graph autoencoder,” *ACM Transactions on Software Engineering and Methodology*, vol. 34, no. 2, pp. 1–28, 2025.
- [15] Z. Zhu, C. Lee, X. Tang, and P. He, “Hemirca: Fine-grained root cause analysis for microservices with heterogeneous data sources,” *ACM Transactions on Software Engineering and Methodology*, vol. 33, no. 8, pp. 1–25, 2024.
- [16] Z. Xie, S. Zhang, Y. Geng, Y. Zhang, M. Ma, X. Nie, Z. Yao, L. Xu, Y. Sun, W. Li *et al.*, “Microservice root cause analysis with limited observability through intervention recognition in the latent space,” in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 6049–6060.
- [17] T. Wang, G. Qi, and T. Wu, “Kgroot: A knowledge graph-enhanced method for root cause analysis,” *Expert Systems with Applications*, vol. 255, p. 124679, 2024.
- [18] L. Zhang, T. Jia, X. Tan, X. Huang, M. Jia, H. Liu, Z. Wu, and Y. Li, “E-log: Fine-grained elastic log-based anomaly detection and diagnosis for databases,” *IEEE Transactions on Services Computing*, 2025.
- [19] J. Lin, P. Chen, and Z. Zheng, “Microscope: Pinpoint performance issues with causal graphs in micro-service environments,” in *Service-Oriented Computing: 16th International Conference, ICSOC 2018, Hangzhou, China, November 12–15, 2018, Proceedings 16*. Springer, 2018, pp. 3–20.
- [20] M. Li, Z. Li, K. Yin, X. Nie, W. Zhang, K. Sui, and D. Pei, “Causal inference-based root cause analysis for online service systems with intervention recognition,” in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3230–3240.
- [21] C.-M. Lin, C. Chang, W.-Y. Wang, K.-D. Wang, and W.-C. Peng, “Root cause analysis in microservice using neural granger causal discovery,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 1, 2024, pp. 206–213.
- [22] G. Yu, P. Chen, H. Chen, Z. Guan, Z. Huang, L. Jing, T. Weng, X. Sun, and X. Li, “Microrank: End-to-end latency issue localization with extended spectrum analysis in microservice environments,” in *Proceedings of the Web Conference 2021*, 2021, pp. 3087–3098.
- [23] G. Yu, Z. Huang, and P. Chen, “Tracerank: Abnormal service localization with dis-aggregated end-to-end tracing data in cloud native systems,” *Journal of Software: Evolution and Process*, vol. 35, no. 10, p. e2413, 2023.
- [24] Z. Zhang, M. K. Ramanathan, P. Raj, A. Parwal, T. Sherwood, and M. Chabbé, “[CRISP]: Critical path analysis of {Large-Scale} microservice architectures,” in *2022 USENIX Annual Technical Conference (USENIX ATC 22)*, 2022, pp. 655–672.
- [25] C. Zhang, Z. Dong, X. Peng, B. Zhang, and M. Chen, “Trace-based multi-dimensional root cause localization of performance issues in microservice systems,” in *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–12.
- [26] W. Zhang, H. Guo, J. Yang, Z. Tian, Y. Zhang, Y. Chaoran, Z. Li, T. Li, X. Shi, L. Zheng *et al.*, “mabc: Multi-agent blockchain-inspired collaboration for root cause analysis in micro-services architecture,” in *Findings of the Association for Computational Linguistics: EMNLP 2024*, 2024, pp. 4017–4033.
- [27] Z. Wang, Z. Liu, Y. Zhang, A. Zhong, J. Wang, F. Yin, L. Fan, L. Wu, and Q. Wen, “Ragent: Cloud root cause analysis by autonomous agents with tool-augmented large language models,” in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024, pp. 4966–4974.
- [28] K. Wang, Y. Li, C. Wang, T. Jia, K. Chow, Y. Wen, Y. Dou, G. Xu, C. Hou, J. Yao *et al.*, “Characterizing job microarchitectural profiles at scale: Dataset and analysis,” in *Proceedings of the 51st International Conference on Parallel Processing*, 2022, pp. 1–11.
- [29] Y. Yang, L. Wang, J. Gu, and Y. Li, “Capturing request execution path for understanding service behavior and detecting anomalies without code instrumentation,” *IEEE Transactions on Services Computing*, vol. 16, no. 2, pp. 996–1010, 2022.
- [30] J. Shen, H. Zhang, Y. Xiang, X. Shi, X. Li, Y. Shen, Z. Zhang, Y. Wu, X. Yin, J. Wang *et al.*, “Network-centric distributed tracing with deepflow: Troubleshooting your microservices in zero code,” in *Proceedings of the ACM SIGCOMM 2023 Conference*, 2023, pp. 420–437.
- [31] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, “Chain-of-thought prompting elicits reasoning in large language models,” *Advances in neural information processing systems*, vol. 35, pp. 24 824–24 837, 2022.
- [32] B. Li, X. Peng, Q. Xiang, H. Wang, T. Xie, J. Sun, and X. Liu, “Enjoy your observability: an industrial survey of microservice tracing and analysis,” *Empirical Software Engineering*, vol. 27, pp. 1–28, 2022.
- [33] S. Luo, H. Xu, C. Lu, K. Ye, G. Xu, L. Zhang, Y. Ding, J. He, and C. Xu, “Characterizing microservice dependency and performance: Alibaba trace analysis,” in *Proceedings of the ACM symposium on cloud computing*, 2021, pp. 412–426.
- [34] X. Zhou, X. Peng, T. Xie, J. Sun, C. Ji, D. Liu, Q. Xiang, and C. He, “Latent error prediction and fault localization for microservice

- applications by learning from system trace logs,” in *Proceedings of the 2019 27th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering*, 2019, pp. 683–694.
- [35] Y. Gan, Y. Zhang, K. Hu, D. Cheng, Y. He, M. Pancholi, and C. Delimitrou, “Seer: Leveraging big data to navigate the complexity of performance debugging in cloud microservices,” in *Proceedings of the twenty-fourth international conference on architectural support for programming languages and operating systems*, 2019, pp. 19–33.
- [36] P. Liu, H. Xu, Q. Ouyang, R. Jiao, Z. Chen, S. Zhang, J. Yang, L. Mo, J. Zeng, W. Xue et al., “Unsupervised detection of microservice trace anomalies through service-level deep bayesian networks,” in *2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2020, pp. 48–58.
- [37] Y. Gan, M. Liang, S. Dev, D. Lo, and C. Delimitrou, “Sage: practical and scalable ml-driven performance debugging in microservices,” in *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, 2021, pp. 135–151.
- [38] Z. Li, J. Chen, R. Jiao, N. Zhao, Z. Wang, S. Zhang, Y. Wu, L. Jiang, L. Yan, Z. Wang et al., “Practical root cause localization for microservice systems via trace analysis,” in *2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS)*. IEEE, 2021, pp. 1–10.
- [39] K. Rasul, A. Ashok, A. R. Williams, A. Khorasani, G. Adamopoulos, R. Bhagwatkar, M. Biloš, H. Ghonia, N. Hassen, A. Schneider et al., “Lag-llama: Towards foundation models for time series forecasting,” in *RO-FoMo: Robustness of Few-shot and Zero-shot Learning in Large Foundation Models*, 2023.
- [40] Y. Liu, H. Zhang, C. Li, X. Huang, J. Wang, and M. Long, “Timer: generative pre-trained transformers are large time series models,” in *Proceedings of the 41st International Conference on Machine Learning*, 2024, pp. 32369–32399.
- [41] A. Das, W. Kong, R. Sen, and Y. Zhou, “A decoder-only foundation model for time-series forecasting,” in *Forty-first International Conference on Machine Learning*, 2024.
- [42] J. Shi, S. Jiang, B. Xu, J. Liang, Y. Xiao, and W. Wang, “Shellgpt: Generative pre-trained transformer model for shell language understanding,” in *2023 IEEE 34th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2023, pp. 671–682.
- [43] S. Liu, D. Yao, L. Fang, Z. Li, W. Li, K. Feng, X. Ji, and J. Bi, “Anomalyilm: Few-shot anomaly edge detection for dynamic graphs using large language models,” *arXiv preprint arXiv:2405.07626*, 2024.
- [44] X. Liu, J. Hu, Y. Li, S. Diao, Y. Liang, B. Hooi, and R. Zimmermann, “Unitime: A language-empowered unified model for cross-domain time series forecasting,” in *Proceedings of the ACM Web Conference 2024*, 2024, pp. 4095–4106.
- [45] H. Guo, J. Yang, J. Liu, L. Yang, L. Chai, J. Bai, J. Peng, X. Hu, C. Chen, D. Zhang et al., “Owl: A large language model for it operations,” *arXiv preprint arXiv:2309.09298*, 2023.
- [46] Y. Liu, Y. Ji, S. Tao, M. He, W. Meng, S. Zhang, Y. Sun, Y. Xie, B. Chen, and H. Yang, “Loglm: From task-based to instruction-based automated log analysis,” *arXiv preprint arXiv:2410.09352*, 2024.
- [47] Y. Chen, H. Xie, M. Ma, Y. Kang, X. Gao, L. Shi, Y. Cao, X. Gao, H. Fan, M. Wen et al., “Automatic root cause analysis via large language models for cloud incidents,” in *Proceedings of the Nineteenth European Conference on Computer Systems*, 2024, pp. 674–688.
- [48] Y. Jiang, C. Zhang, S. He, Z. Yang, M. Ma, S. Qin, Y. Kang, Y. Dang, S. Rajmohan, Q. Lin et al., “Xpert: Empowering incident management with query recommendations via large language models,” in *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, 2024, pp. 1–13.
- [49] D. Zhang, X. Zhang, C. Bansal, P. Las-Casas, R. Fonseca, and S. Rajmohan, “Lm-pace: Confidence estimation by large language models for effective root causing of cloud incidents,” in *Companion Proceedings of the 32nd ACM International Conference on the Foundations of Software Engineering*, 2024, pp. 388–398.
- [50] P. Hamadanian, B. Arzani, S. Fouladi, S. K. R. Kakarla, R. Fonseca, D. Billor, A. Cheema, E. Nkposong, and R. Chandra, “A holistic view of ai-driven network incident management,” in *Proceedings of the 22nd ACM Workshop on Hot Topics in Networks*, 2023, pp. 180–188.
- [51] J. Pan, W. S. Liang, and Y. Yidi, “Raglog: Log anomaly detection using retrieval augmented generation,” in *2024 IEEE World Forum on Public Safety Technology (WFPST)*. IEEE, 2024, pp. 169–174.
- [52] W. Zhang, Q. Zhang, E. Yu, Y. Ren, Y. Meng, M. Qiu, and J. Wang, “Lograg: Semi-supervised log-based anomaly detection with retrieval-augmented generation,” in *2024 IEEE International Conference on Web Services (ICWS)*. IEEE, 2024, pp. 1100–1102.
- [53] L. Zhang, T. Jia, M. Jia, Y. Wu, H. Liu, and Y. Li, “Xraglog: A resource-efficient and context-aware log-based anomaly detection method using retrieval-augmented generation,” in *AAAI 2025 Workshop on Preventing and Detecting LLM Misinformation (PDLIM)*, 2025.
- [54] ———, “Scalog: Scalable log-based failure diagnosis using llm,” in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.
- [55] L. Zhang, Y. Zhai, T. Jia, X. Huang, C. Duan, and Y. Li, “Agentfm: Role-aware failure management for distributed databases with llm-driven multi-agents,” *arXiv preprint arXiv:2504.06614*, 2025.
- [56] L. Zhang, Y. Zhai, T. Jia, C. Duan, S. Yu, J. Gao, B. Ding, Z. Wu, and Y. Li, “Thinkfl: Self-refining failure localization for microservice systems via reinforcement fine-tuning,” *arXiv preprint arXiv:2504.18776*, 2025.
- [57] L. Zhang, T. Jia, M. Jia, and Y. Li, “Logdb: Multivariate log-based failure diagnosis for distributed databases (extended from multilog),” *arXiv preprint arXiv:2505.01676*, 2025.
- [58] C. Duan, T. Jia, Y. Yang, G. Liu, J. Liu, H. Zhang, Q. Zhou, Y. Li, and G. Huang, “Eagerlog: Active learning enhanced retrieval augmented generation for log-based anomaly detection,” in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.
- [59] M. He, T. Jia, C. Duan, H. Cai, Y. Li, and G. Huang, “Weakly-supervised log-based anomaly detection with inexact labels via multi-instance learning,” in *2025 IEEE/ACM 47th International Conference on Software Engineering (ICSE)*. IEEE Computer Society, 2025, pp. 726–726.
- [60] C. Duan, Y. Yang, T. Jia, G. Liu, J. Liu, H. Zhang, Q. Zhou, Y. Li, and G. Huang, “Famos: Fault diagnosis for microservice systems through effective multi-modal data fusion,” in *2025 IEEE/ACM 47th International Conference on Software Engineering (ICSE)*. IEEE Computer Society, 2025, pp. 610–610.
- [61] P. Xiao, T. Jia, C. Duan, H. Cai, Y. Li, and G. Huang, “Logcae: An approach for log-based anomaly detection with active learning and contrastive learning,” in *2024 IEEE 35th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2024, pp. 144–155.
- [62] M. He, T. Jia, C. Duan, H. Cai, Y. Li, and G. Huang, “Llmelog: An approach for anomaly detection based on llm-enriched log events,” in *2024 IEEE 35th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2024, pp. 132–143.
- [63] C. Duan, T. Jia, H. Cai, Y. Li, and G. Huang, “Afalog: A general augmentation framework for log-based anomaly detection with active learning,” in *2023 IEEE 34th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2023, pp. 46–56.
- [64] C. Duan, T. Jia, Y. Li, and G. Huang, “Aclog: An approach to detecting anomalies from system logs with active learning,” in *2023 IEEE International Conference on Web Services (ICWS)*. IEEE, 2023, pp. 436–443.
- [65] L. Zhang, L. Fang, C. Duan, M. He, L. Pan, P. Xiao, S. Huang, Y. Zhai, X. Hu, P. S. Yu et al., “A survey on parallel text generation: From parallel decoding to diffusion language models,” *arXiv preprint arXiv:2508.08712*, 2025.