UNIVERSITY OF TORONTO

MIDDLEWARE SYSTEMS RESEARCH GROUP MSRG.ORG

From Signals to Solutions: Grounded AI Agents for Root Cause Analysis in Complex Systems

Hans-Arno Jacobsen, FIEEE, FCAE

Tier 1 Canada Research Chair in Data-Intensive Systems

Jeffrey Skoll Chair in Computer Networks and Innovation

University of Toronto

IEEE International Conference on E-Business Engineering (ICEBE 2026)

Nov 20th, 2026, Wuhan, China

Scan me

UNIVERSITY OF TORONTO

MIDDLEWARE SYSTEMS RESEARCH GROUP MSRG.ORG

From Signals to Solutions: Grounded AI Agents for Root Cause Analysis in Complex Systems

Bell

Hans-Arno Jacobsen, FIEEE, FCAE

Tier 1 Canada Research Chair in Data-Intensive Systems

Jeffrey Skoll Chair in Computer Networks and Innovation

University of Toronto

NSERC CRSNG

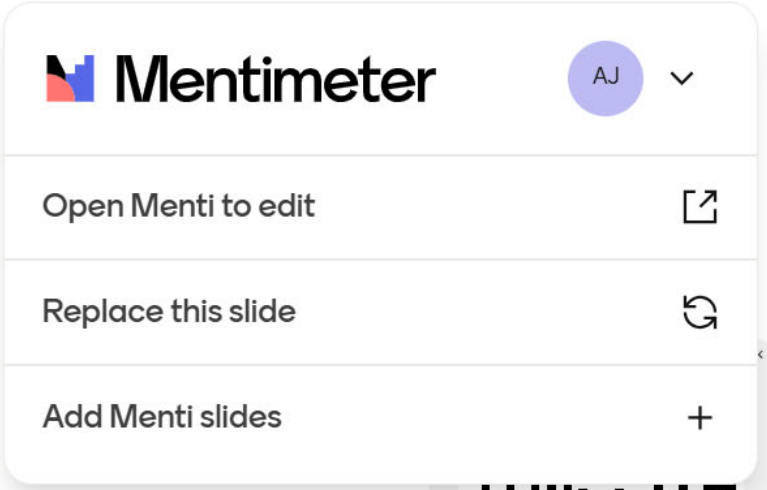
Scan me

Joint work with Yifang Tian, Ph.D. Candidate



MIDDLEWARE SYSTEMS
RESEARCH GROUP
MSRG.ORG





Let's get to know our audience!

Please answer three quick questions.

IEEE ICEBE 2026 · 19–21 November 2026 · Wuhan, China

IEEE International Conference on E-Business Engineering

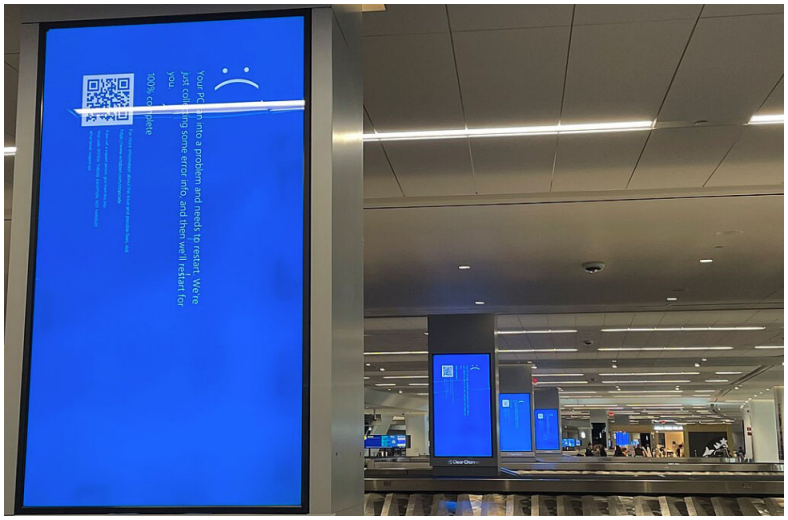


menti.com

8129 2455

12 joined

CrowdStrike Incident



Date: July 19, 2024

Impacted Users:
8.5 million Windows user

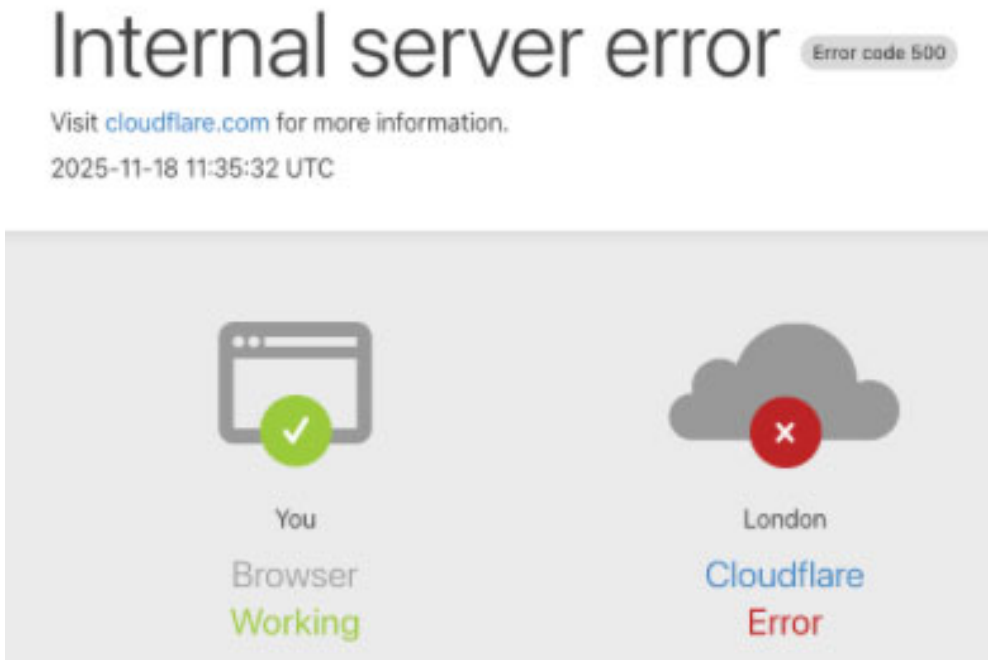
Root Cause: Faulty configuration update

Financial loss: $5.4 billion

CROWDSTRIKE

(c) 2026 - From Signals to Solutions

Cloudflare Incident

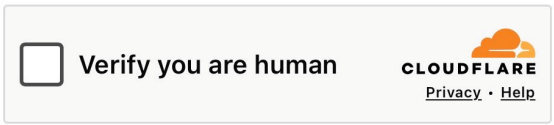


Date: November 18, 2025

Impacted Users:
~2.65 billion

Root Cause: A permissions change in a database triggered a bug

Duration: 6 hours

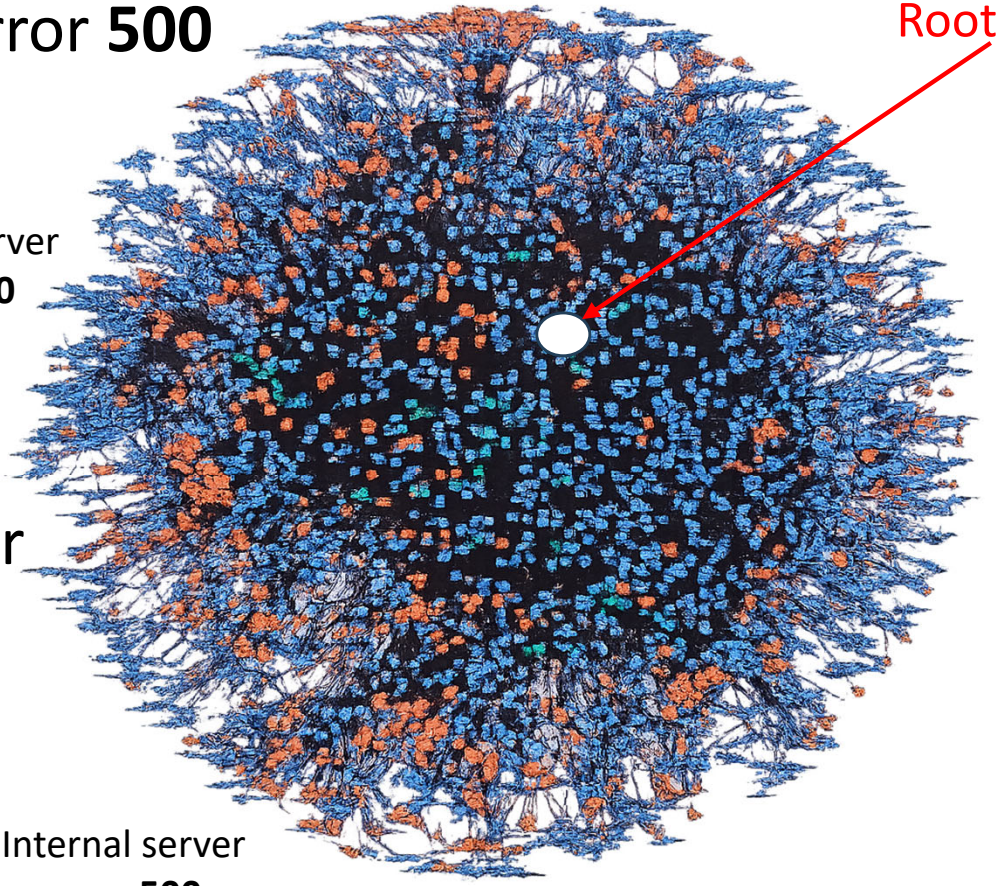


(c) 2026 - From Signals to Solutions

Microservices at Amazon

Ca. 2008

Internal server error **500**



W. Vogels, X, 12:48 PM · Jun 11, 2016

Agenda

- Operational Challenge: *When Microservices Fail*
- Root Cause Analysis (RCA)
- **GALA**: Graph-Augmented LLM Agents for RCA
- **SURE-Score**: A New Evaluation Standard
- Evaluation Results
- Our Research Roadmap



(c) 2026 - From Signals to Solutions

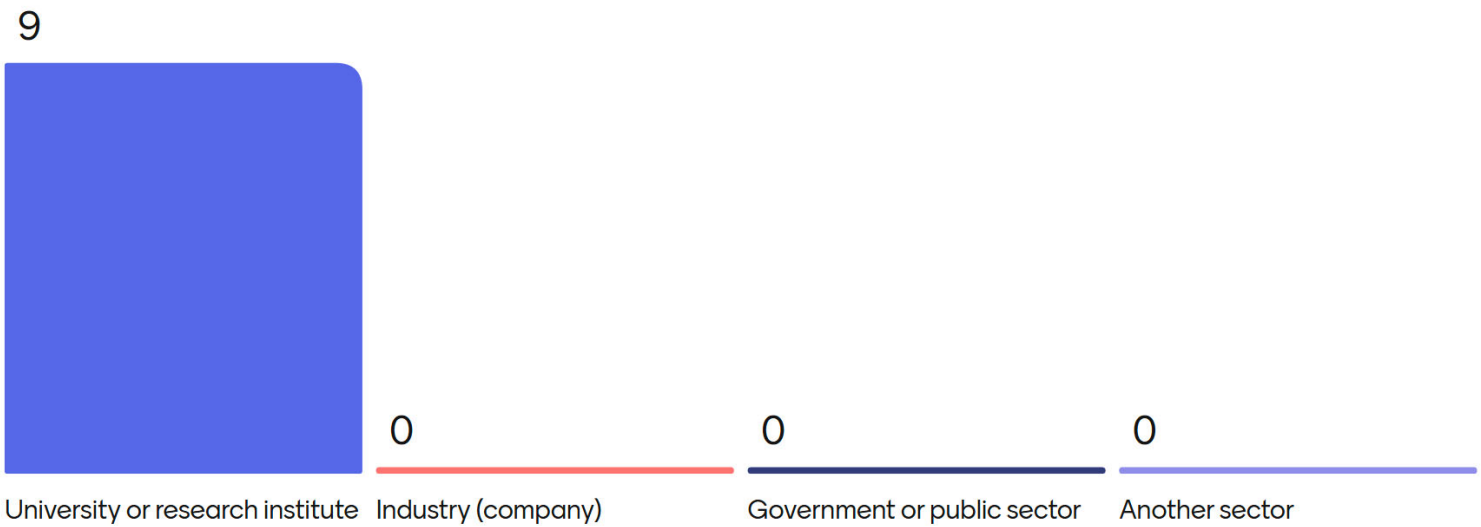
Mentimeter AJ

Open Menti to edit

Replace this slide

Add Menti slides

Which sector are you mainly affiliated with?



menti.com

8129 2455

9 of 12 responded



Root Cause

Root Cause Analysis: **Systematic Identification of Root Cause**

- **Fault**: Underlying *defect* or *abnormal condition* in a system (e.g., software bug, misconfiguration, hardware defect)

- **Failure:** Dynamic, active event where the system does not meet its specifications

- **Root cause:**
  - Fundamental, underlying fault or condition responsible for an observed failure.
  - Addressing it prevents recurrence of the same failure mechanism.

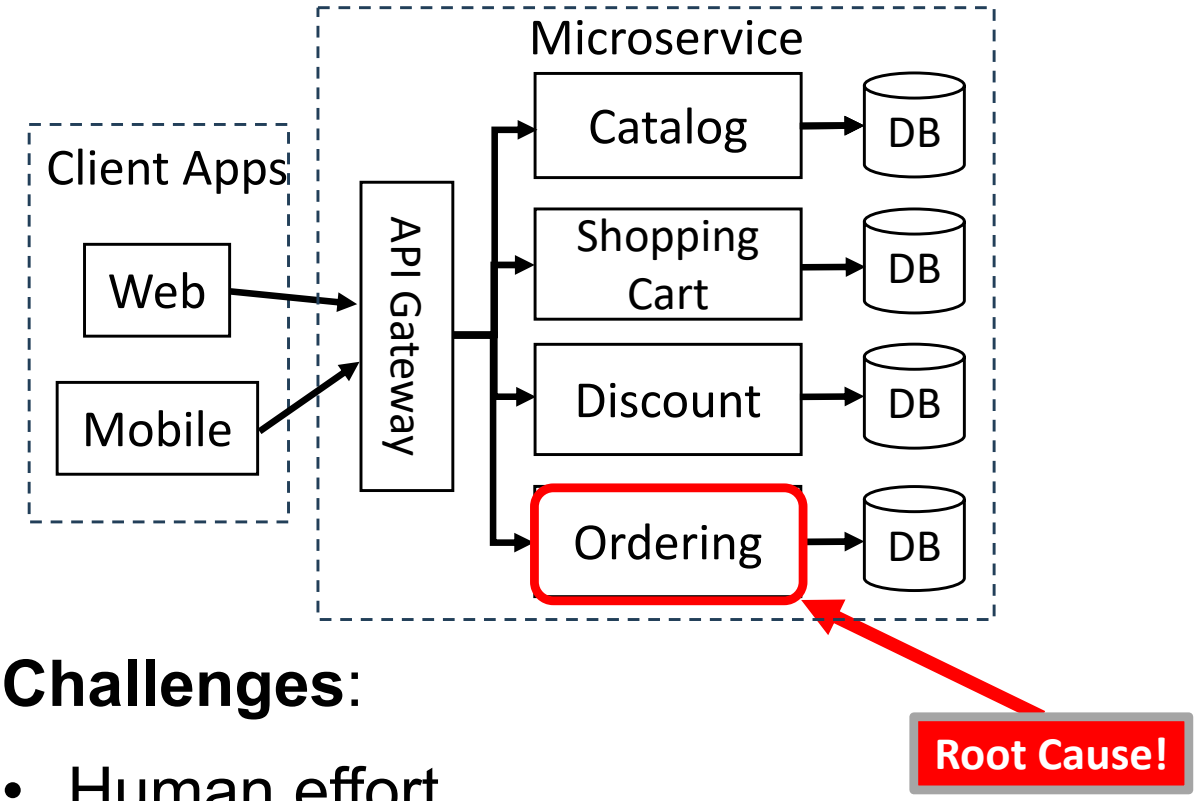
(c) 2026 - From Signals to Solutions

Why Root Cause Analysis Matters



Root Cause Analysis (RCA)

Targeted at Microservices Architectures



Challenges:

- Human effort
- Multiple data modalities
- Large volume of data

Objectives:

- Bridge **statistical causality** and **LLM reasoning**
- Deliver **interpretable**, **multi-modal diagnostics**
- Set a new **standard for RCA evaluation**

(c) 2026 - From Signals to Solutions

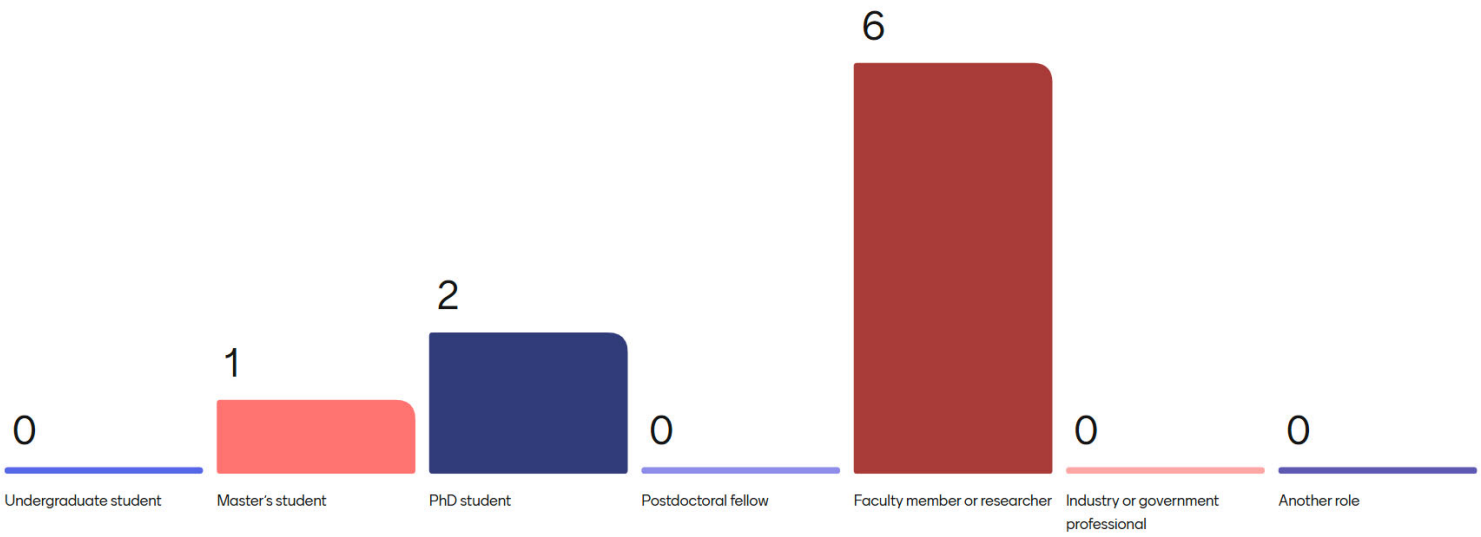
Mentimeter AJ

Open Menti to edit

Replace this slide

Add Menti slides

Which best describes your current role?



menti.com

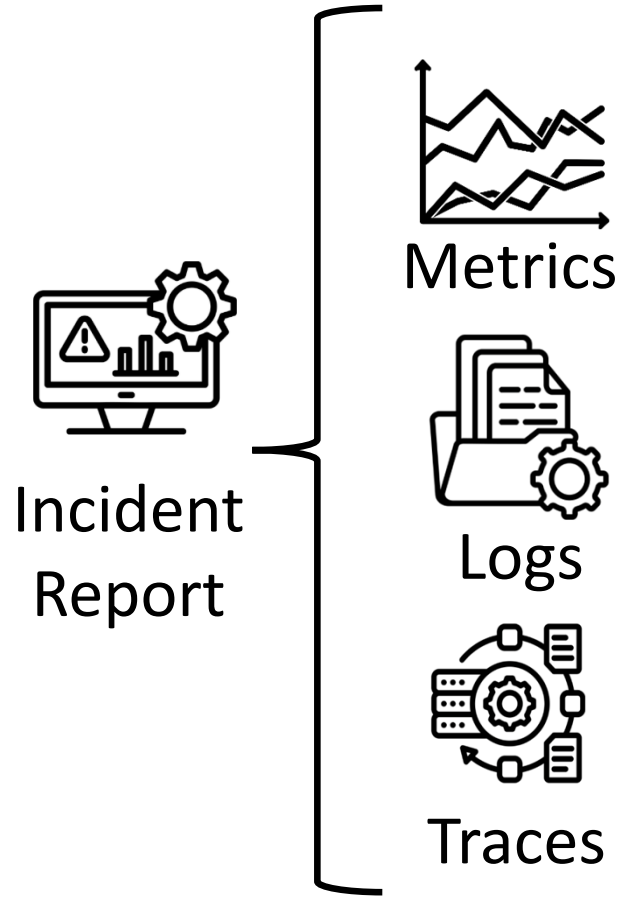
8129 2455

9 of 12 responded

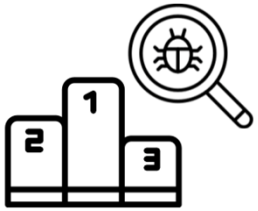
Goal:

Automate RCA for Microservices Architectures

Input:



Expected output:



Final Root Cause Ranking



Incident Summarization



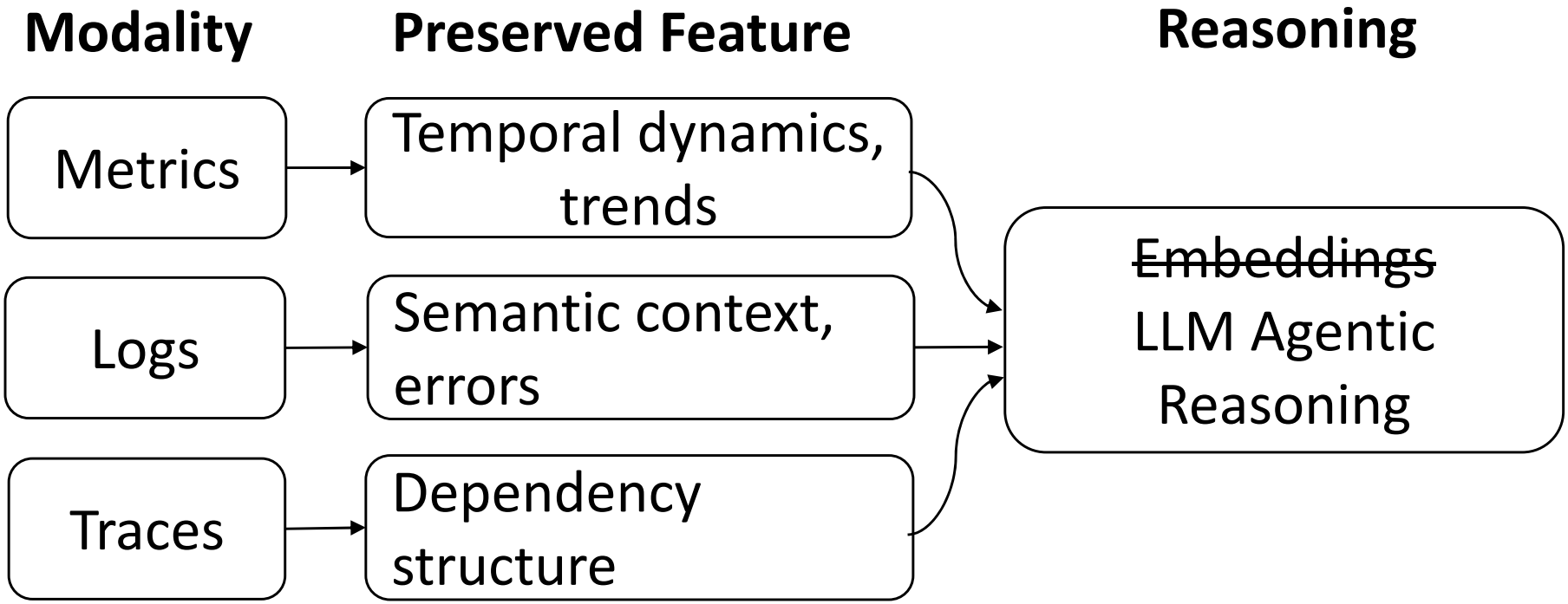
Action Recommendation

(c) 2026 - From Signals to Solutions

Design Principle I

Modality-Preserving Multi-Modal Data Integration

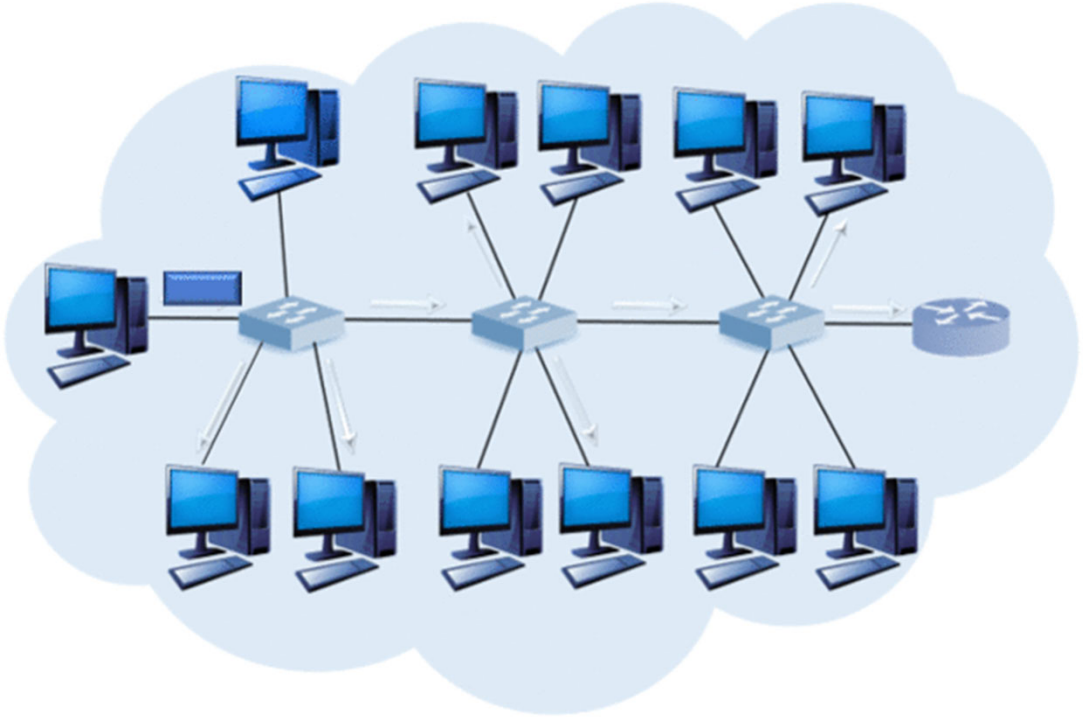
Fusion is not homogenization: it's coordination.



Design Principle II

Scalable Graph- and Time-Aware Causal Reasoning

Integrate **graph topology** and **temporal causality** to model fault propagation paths.



https://commons.wikimedia.org/wiki/File:GIF_Broadcast_traffic.gif

Design Principle III

Improving RCA via Iterative Agentic Reasoning

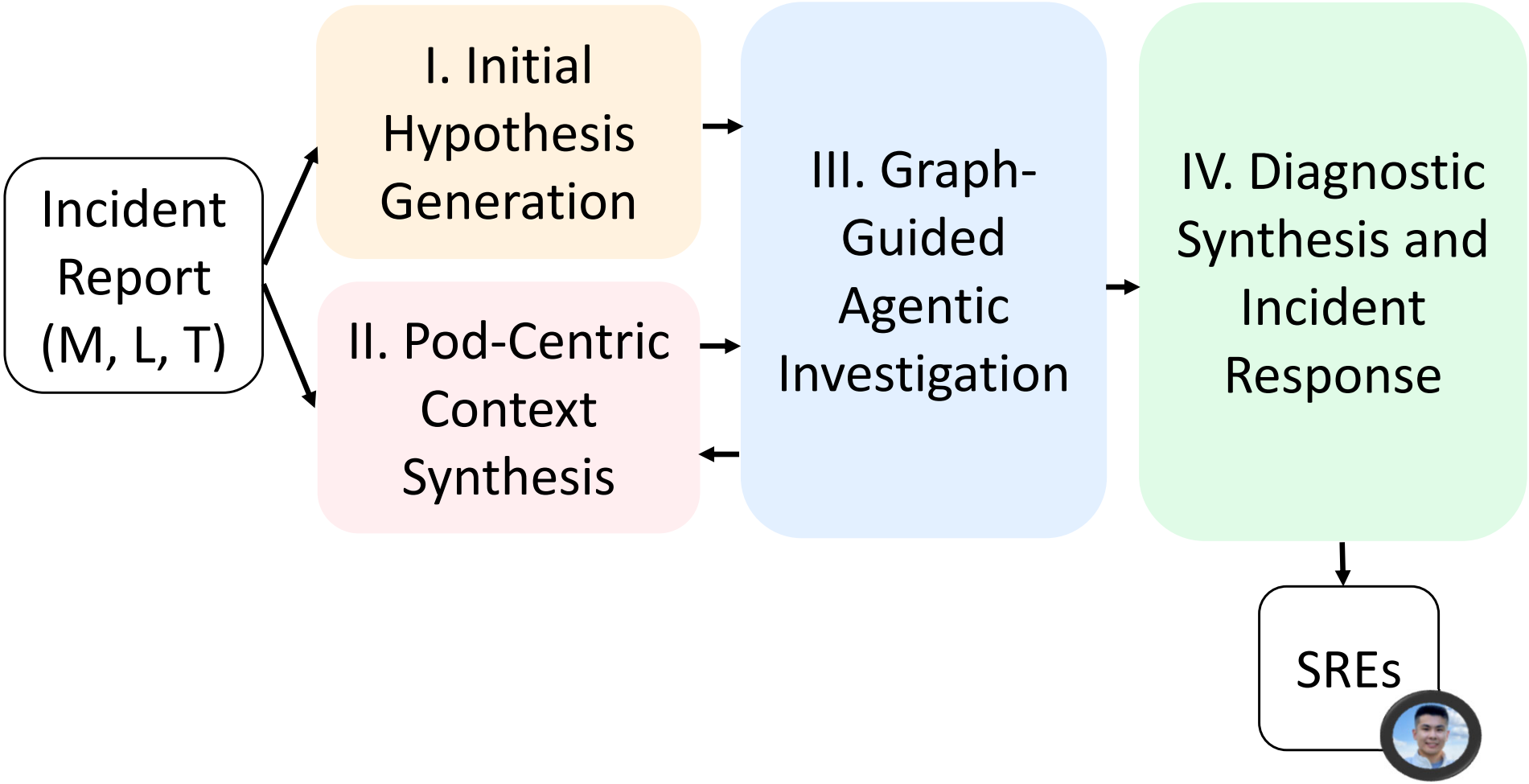
- RCA approaches stop at ranking; they don't "reason"
- Instead, iteratively plan, analyze, and re-evaluate evidence
- Ground reasoning flow in service dependency to reduce hallucination



(c) 2026 - From Signals to Solutions

GALA Architecture

Graph-**A**ugmented Large **L**anguage Model **A**gentic Workflow – **Our Approach**

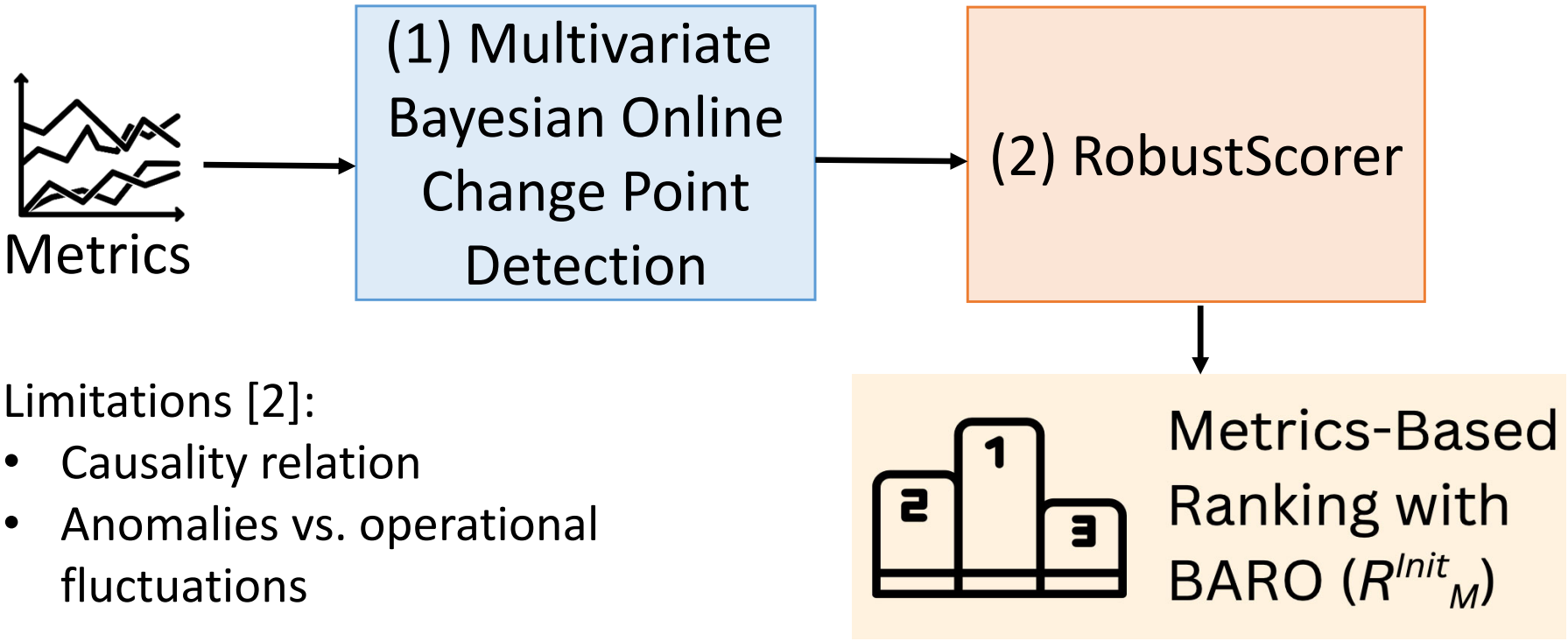


(c) 2026 - From Signals to Solutions

Initial Hypothesis Generation I

Temporal Causality

We employ BARO[1] , a non-universal[2] algorithm



Limitations [2]:

- Causality relation
- Anomalies vs. operational fluctuations

[1] Pham, L., Ha, H., & Zhang, H. (2024). BARO: Robust root cause analysis for microservices via multivariate Bayesian online change point detection. Proceedings of the ACM on Software Engineering, (FSE), 2214-2237.
[2] Pham, L., Ha, H., & Zhang, H. (2024). Root Cause Analysis for Microservices based on Causal Inference: How Far Are We?. In *2024 39th IEEE/ACM International Conference on Automated Software Engineering (ASE)* (pp. 706-718).

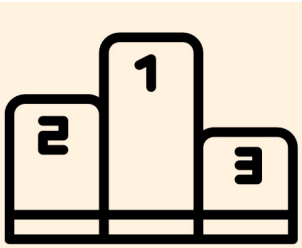
Initial Hypothesis Generation II

Structural **T**race **R**anking via **I**ntermodal e**X**ploration

Structural causality with *STRIX*

Construct service dependency graph and rank suspicious services along three dimensions:





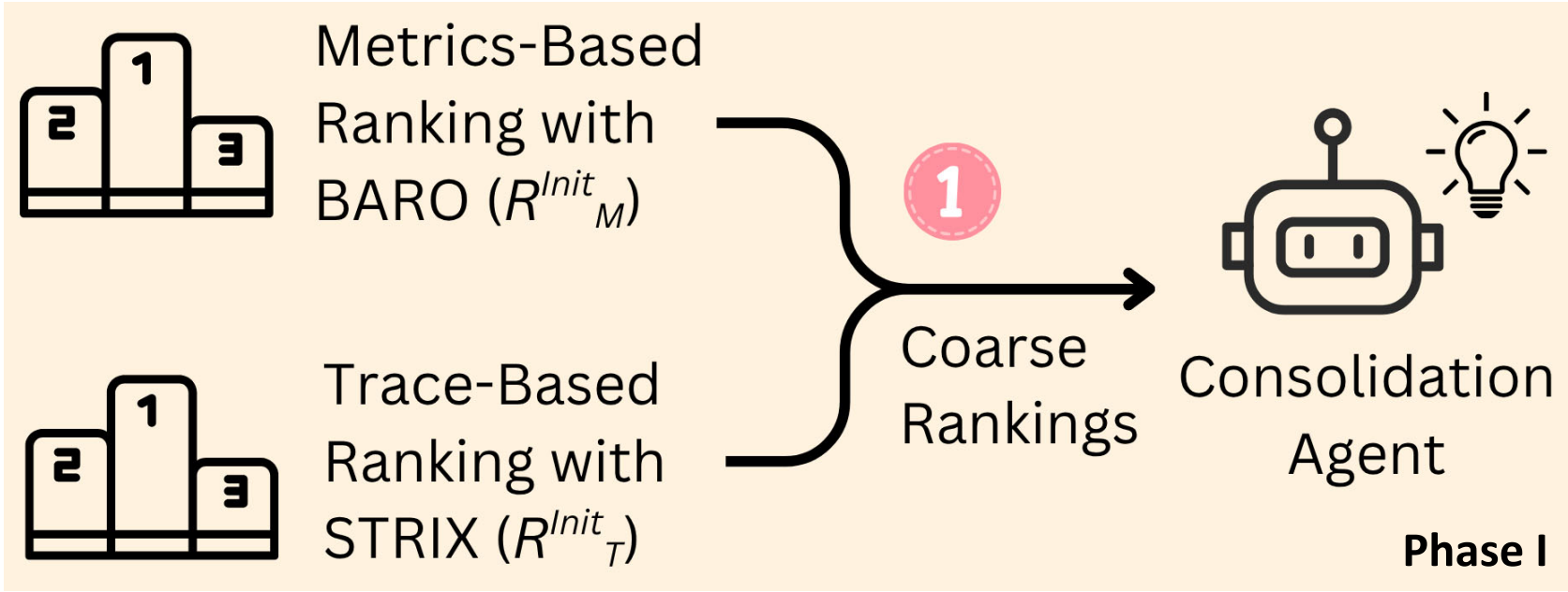
Trace-Based Ranking with STRIX (R^{Init}_{T})

(c) 2026 - From Signals to Solutions

Initial Hypothesis Generation III

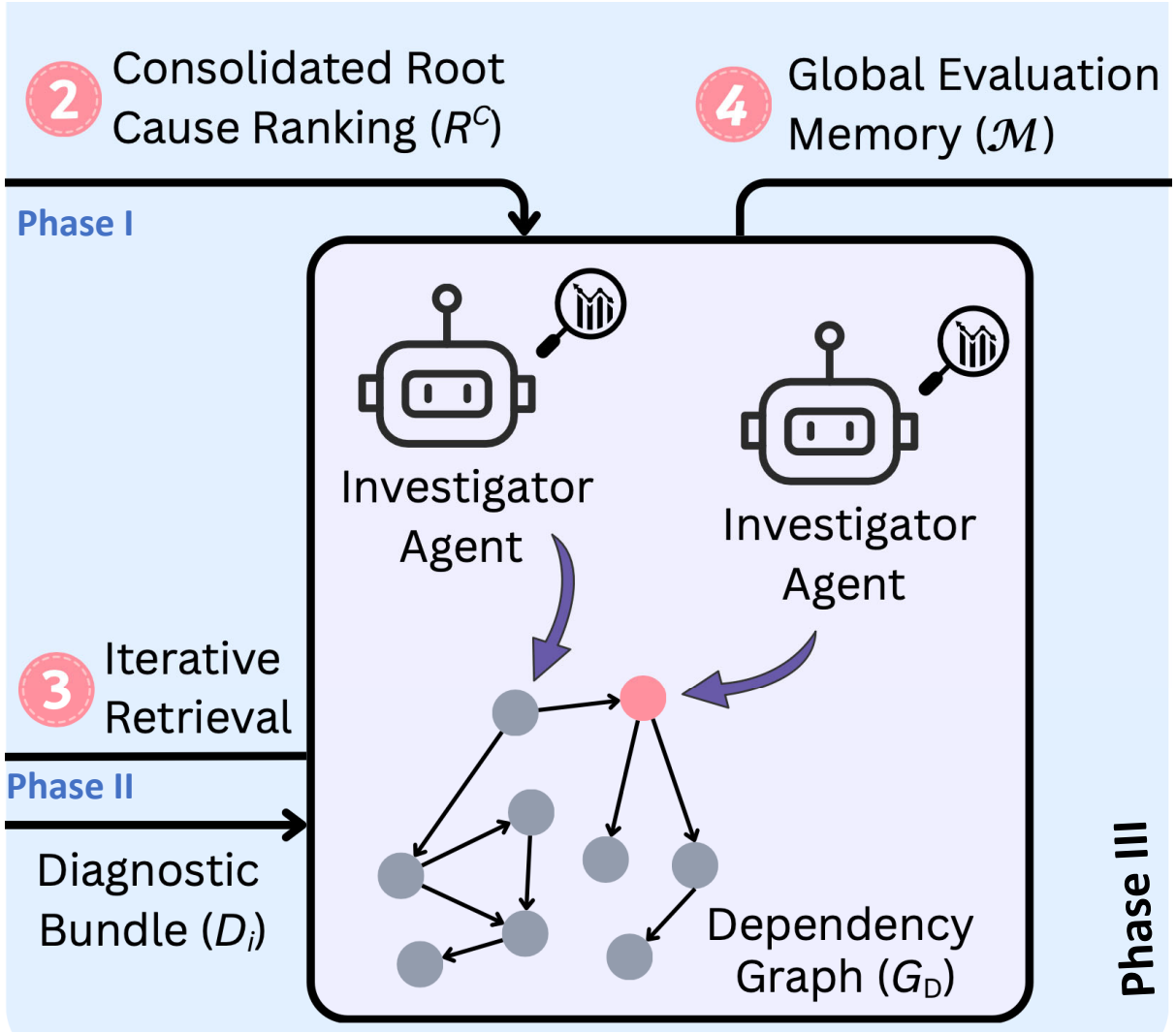
Consolidation Agent

- **Input**: Temporal, structural coarse ranking
- **Cross-modal consensus**: higher place in ranking
- Produce consolidated ranking and confidence with explicit evidence provenance



Agentic Investigation

Graph-based Exploration



Agents follow fault-propagation edges with confidence-driven, depth-bounded expansion, substantially reducing the search space.

(c) 2026 - From Signals to Solutions

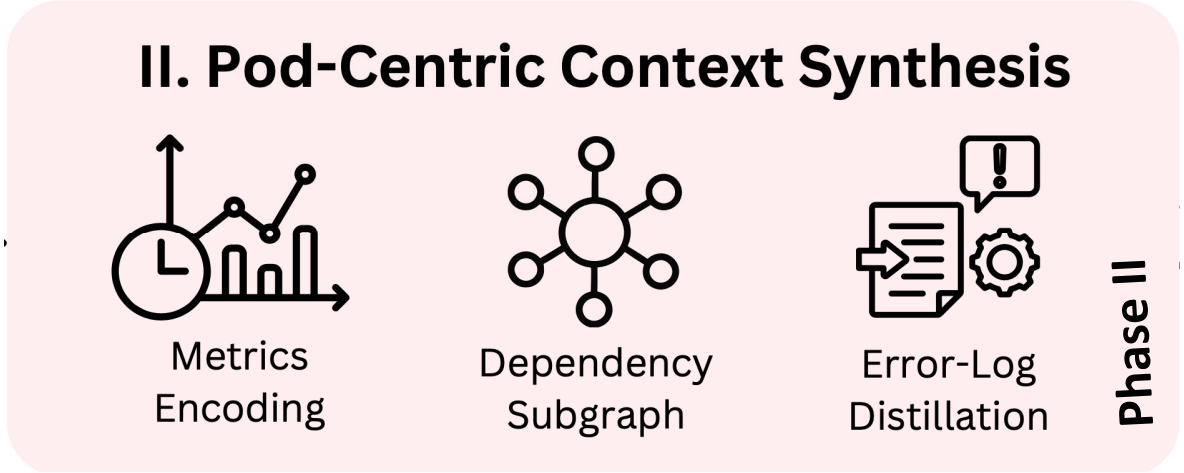
Agentic Reasoning & Re-ranking

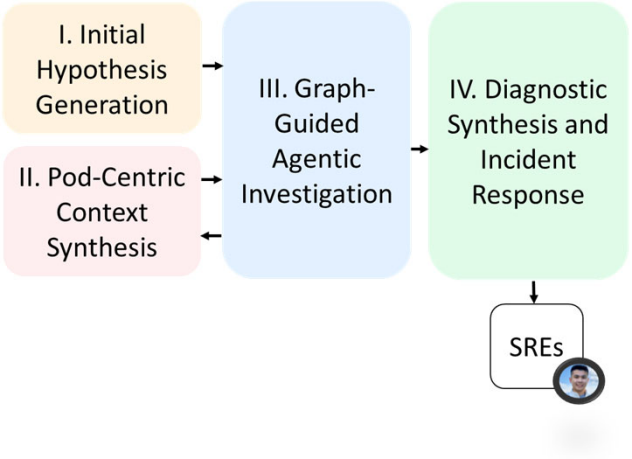
Investigator Agent Design

In topology-guided parallel search, Investigator Agent inspects each node's ***Diagnostic Bundle***:

- **low confidence** expands to neighbours
- **high confidence** halts and commits to "**global memory**"



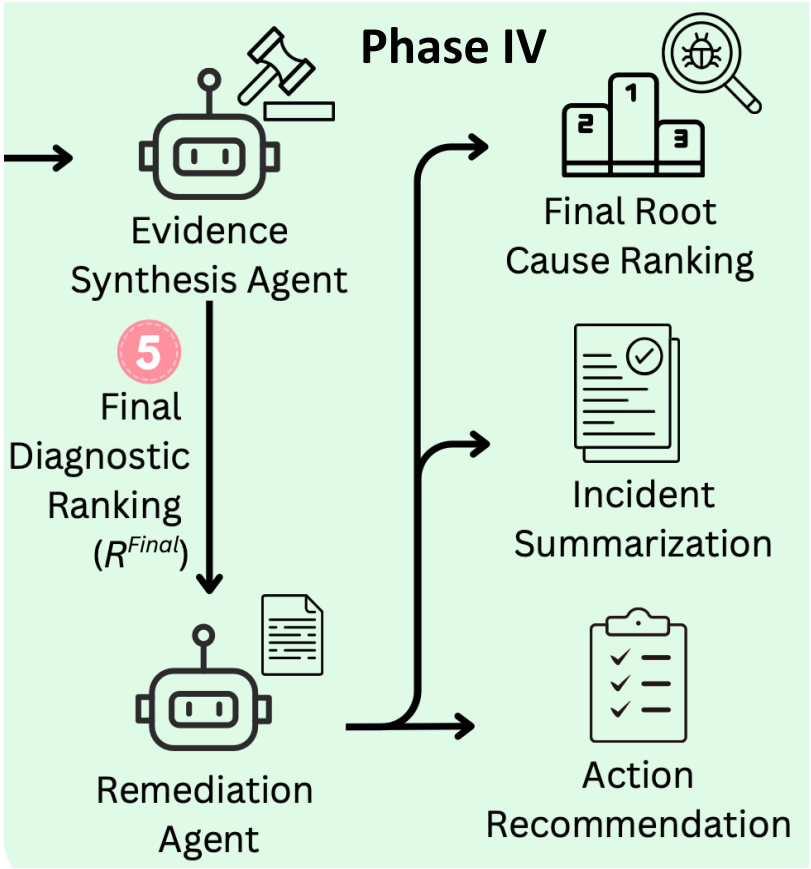




(c) 2026 - From Signals to Solutions

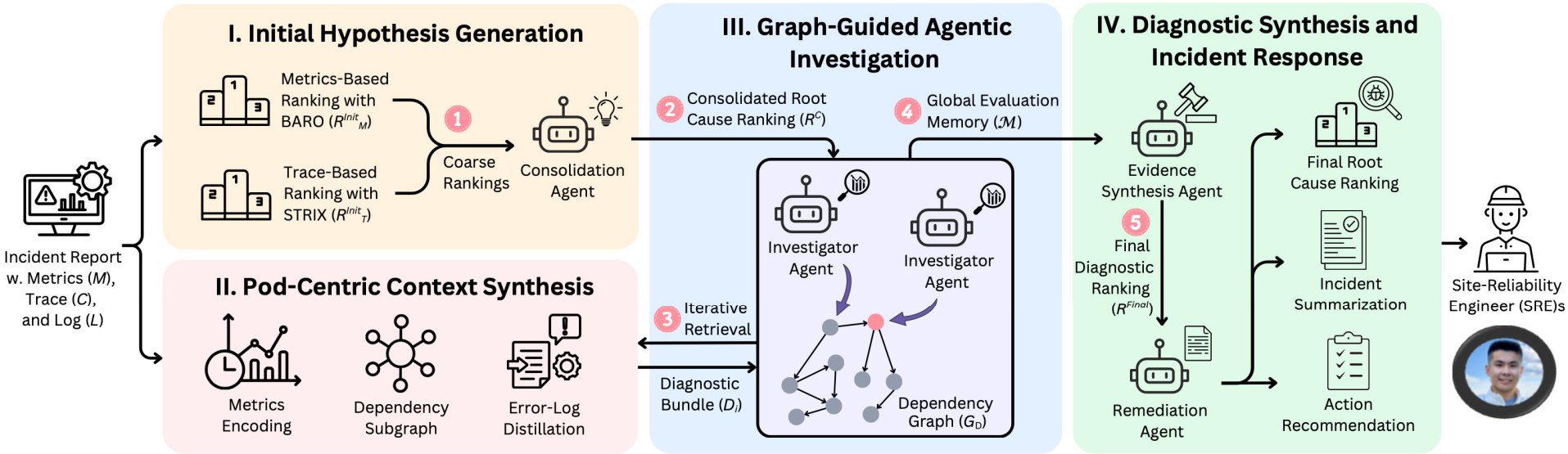
Diagnostic Synthesis and Incident Response

After ***Graph-Guided Agentic Investigation***, all LLM trajectories from global memory feed into the ***Evidence Synthesis Agent***.



GALA Architecture

Graph-**A**ugmented Large **L**anguage Model **A**gentic Workflow



Mentimeter AJ

Open Menti to edit

Replace this slide

Add Menti slides

What is your main area of expertise? Please enter 1–3 words in English. Examples: AI/ML, cloud computing, networking, software engineering, data management.



menti.com
8129 2455

9 of 12 responded

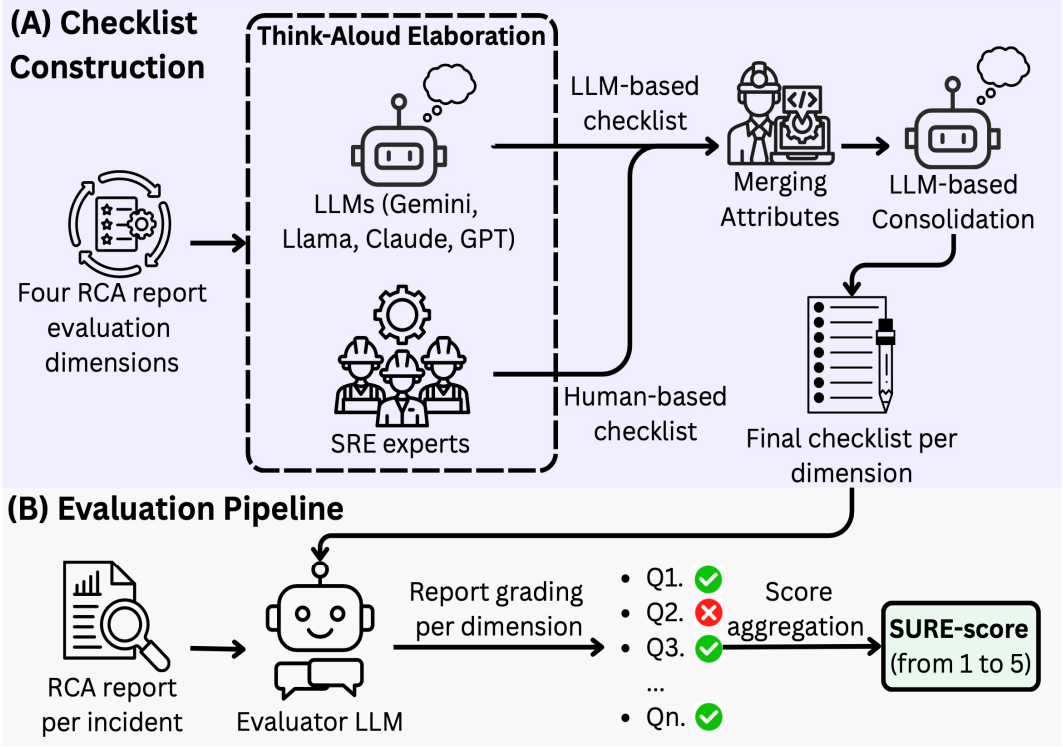
SURE-Score for Evaluation

SUmmarization **RE**commendation Score

> **Limitations for AIOps evaluations:** Traditional scores (BLEU, ROUGE, BERTScore) measure text similarity, not RCA-specific quality.

SURE-Score Dimensions

- Evidence Groundedness
- Operational Feasibility
- Diagnostic Precision
- Triage Logic





Evaluation

Research Questions

- Root cause ranking effectiveness
- Incident summarization, action recommendation effectiveness
- Effect of LLM choice on performance, efficiency, and cost

- Contribution of key components to performance
- Effect of agentic investigation parameters

(c) 2026 - From Signals to Solutions

Evaluation

Setup - Metrics

- **Microservices:** OnlineBoutique and TrainTicket
  - Comprise 12 and 69 services, respectively
  - 90 failure cases and 6 failure types each
  - Millions of telemetry data records

- **Metrics:**
  - AC@1, AC@3, MRR (mean reciprocal rank)
  - Text Quality: Traditional (BERTScore, Cosine Similarity) and our **SURE-Score**

- Ubuntu 22.04 VM 8 cores, 16 GB RAM, 500 GB SSD

(c) 2026 - From Signals to Solutions

Evaluation

Baselines

- **LLMs:**
  - **Proprietary**: **gemini**-2.5-flash, **claude**-sonnet-4-5, **GPT**-4.1-mini, and **GPT**-4.1
  - **Open source**: **deepseek-**v3., **Qwen3**-Next-80B-A3B-Instruct, and **llama**-4-maverick
- **Non LLM-based:** BARO, Granger, CausalRCA, PC, LiNGAM, TraceRCA and Nezha
- **LLM-based:** mABC, Flow-of-Action, and RCA-Agent

(c) 2026 - From Signals to Solutions

Evaluation Result

GALA Effectiveness

Type	Method	Online Boutique					Train Ticket				
		AC@1 (%)	AC@3 (%)	MRR	Time (s)	$\text{Token}_{\text{total}}$	AC@1 (%)	AC@3 (%)	MRR	Time (s)	$\text{Token}_{\text{total}}$
Non-LLM	BARO	14.44	87.78	0.524	**0.05**	-	66.67	82.22	0.759	**0.27**	-
	CausalRCA	26.67	68.89	0.490	1,575	-	-	-	-	T/O	-
	PC-PR	15.56	36.67	0.332	6.68	-	1.11	6.67	0.082	1,746	-
	PC-RW	3.33	43.33	0.290	6.57	-	1.11	4.44	0.077	1,799	-
	LiNGAM-PR	0.00	1.11	0.086	0.50	-	0.00	0.00	0.021	48.90	-
	LiNGAM-RW	3.33	43.33	0.292	0.69	-	1.11	4.44	0.079	60.53	-
	Granger-PR	12.22	35.56	0.313	10.86	-	3.33	6.67	0.087	400	-
	Granger-RW	3.33	44.44	0.293	11.15	-	1.11	3.33	0.076	406	-
	TraceRCA	13.19	61.54	0.340	3.23	-	64.44	75.56	0.711	13.45	-
	Nezha	28.89	52.22	0.436	76.21	-	15.56	40.00	0.324	513	-
LLM-based	GraphRAG	21.35	35.96	0.331	9.58	78.2K	7.78	14.44	0.163	19.62	149.3K
	ReAct	44.44	65.56	0.549	14.73	12.6K	24.44	41.39	0.353	15.92	28.3K
	mABC	13.33	44.44	0.296	57.02	21.9K	12.22	23.33	0.196	92.98	38.9K
	RCA-Agent	20.00	48.89	0.337	50.63	60.0K	32.22	58.89	0.440	89.81	116.4K
	Flow-of-Action	48.35	70.33	0.590	36.50	**9.6K**	47.78	74.44	0.627	123	**24.5K**
	GALA	**74.44**	**98.89**	**0.854**	52.87	32.7K	**73.33**	**85.56**	**0.801**	63.16	90.4K
	Improvement	+54%	+13%	+45%	-	-	+10%	+4%	+6%	-	-

> GALA exhibits strongest root-cause ranking performance with runtime and token consumption trade-off.

Evaluation Result

Incident Summarization and Action Recommendation

Models: Q = Qwen3, L = Llama-4
Traditional: BE = BERTScore, CS = Cosine Similarity

		Trad.		SURE-Score (Q)					SURE-Score (L)				
	Method	BE	CS	E	O	D	T	F	E	O	D	T	F
OB	BARO+LLM	-.10	.28	1.0	1.4	1.7	1.3	1.3	1.0	3.4	2.2	2.2	2.2
	mABC	-.23	.21	1.0	1.0	1.8	1.4	1.3	1.3	1.0	2.6	3.7	2.2
	Flow-of-A.	-.23	.19	1.4	1.0	2.3	1.0	1.4	1.7	1.0	2.4	1.2	1.6
	RCA-Agent	-.44	.08	1.0	1.0	1.6	1.0	1.2	1.0	1.0	1.2	1.0	1.0
	GALA	**.04**	**.35**	**1.8**	**3.5**	**4.4**	**3.3**	**3.3**	**5.0**	**3.7**	**4.5**	**5.0**	**4.5**
TT	BARO+LLM	-.06	.37	1.0	1.4	2.1	1.4	1.5	1.0	3.1	2.5	1.9	2.1
	mABC	-.17	.25	1.0	1.0	2.1	1.4	1.4	1.2	1.0	2.8	3.1	2.0
	Flow-of-A.	-.25	.23	1.3	1.0	2.8	1.0	1.5	2.3	1.0	2.8	1.2	1.8
	RCA-Agent	-.41	.14	1.0	1.0	1.8	1.0	1.2	1.0	1.0	1.2	1.0	1.1
	GALA	**.06**	**.38**	**1.4**	**3.3**	**4.2**	**3.2**	**3.0**	**4.9**	**3.5**	**4.4**	**4.8**	**4.4**

SURE-Score
E Evidence Groundedness
O Operational Feasibility
D Diagnostic Precision
T Triage Logic
F Final (mean)

GALA produces **most grounded summaries** and **actionable recommendations** across all LLM backends.

(c) 2026 - From Signals to Solutions



Evaluation Result

Impact of LLM Selection

LLM	AC@1	AC@3	MRR	Token	Time (s)	$
gemini-2.5-flash	55.6	87.8	.718	**17.8K**	**18.5**	.030
deepseek-v3	67.8	94.4	.810	35.2K	64.9	.044
claude-sonnet-4-5	71.1	93.3	.827	45.6K	101.3	.255
GPT-4.1-mini	74.4	**98.9**	.854	32.7K	52.9	.022
Qwen3-80B	76.7	95.6	.855	40.2K	21.3	.015
llama-4-maverick	77.8	97.8	.872	32.0K	37.4	**.003**
gpt-4.1	**78.9**	94.4	**.873**	34.0K	48.4	.110

GALA benefits from stronger LLMs, but **open-source models** can **deliver near-top performance at substantially lower cost.**

(c) 2026 - From Signals to Solutions



Conclusions

Takeaways & Outlook

- **GALA:** RCA framework for multi-modal, graph-guided LLM investigation
- **SURE-Score:** Human-seeded LLM-as-a-judge metric
- GALA automates root cause analysis and actionable incident reporting for on-call engineers (SREs)
- Boost accuracy by leveraging even more modalities
- Reduce token footprint, increase information content
- Tune models based on incident history
- Train SLM vs. use LLMs, accuracy per token?

(c) 2026 - From Signals to Solutions

MIDDLEWARE SYSTEMS RESEARCH GROUP MSRG.ORG

Our AIOps Literature

GALA basis for talk today

- **GALA+**: Graph-Augmented LLM Agents for Root Cause Analysis and Incident Response in Microservices [ASE 2026] (GALA arXiv identifier: 2508.12472)
- **AgentChaosBench**: Benchmarking External Fault Detection and Localization in Agentic Systems [ASE-TRUST 2026]
- When Log Parsing Removes What Matters: Detecting Failures in Distributed GPU Workloads [ASE-RASE 2026]
- **SREGym:** A Live Benchmark for AI SRE Agents with High-Fidelity Failure Scenarios [**NeurIPS 2026**]
- **FAME**: Failure-Aware Mixture-of-Experts for Message-Level Log Anomaly Detection [ISSRE 2026]
- **CERCL**: Context Engineering for Single-Shot Microservice Root Cause Localization [arXiv: 2610."*TBD*"]
- **ORCA**: Observability-Grounded Program Repair for Microservice Incidents [arXiv: 2608.17018]
- **SRE-Marathon**: A Continuous, Change-Driven Benchmark for Autonomous Site Reliability Agents [arXiv: 2609.33023]

GALA paper

SREGym paper

Code & artifacts

https://msrg.github.io/research/rca-aiops/

Research Roadmap

The Incident Lifecycle

Detect

Is something wrong?

- **FAME** — ISSRE 2026
- **GPU-log study** — ASE-RASE 2026
- **AgentChaosBench** — ASE-TRUST 2026

Diagnose

Where did it start?

- **GALA** — ASE 2026
- **CERCL** — under review
- **RECAP** — under review

Repair

How do we fix it?

- **ORCA** — under review
- **OPEN GAP** — ≤ 2.4% of faults repaired while active

Evaluate

Does it hold up in operation?

- **SRE-Marathon** — under review
- **SREGym** — NeurIPS 2026

REF SREGym: Clark et al., NeurIPS 2026, arXiv:2605.07161 · FAME: Wang et al., ISSRE 2026, arXiv:2605.22779 · ORCA: Gao et al., arXiv:2608.17018 · AgentChaosBench: Zhang et al., ASE-TRUST 2026 · GPU logs: Cui et al., ASE-RASE 2026

MIDDLEWARE SYSTEMS
RESEARCH GROUP
MSRG.ORG

MSRG AIOps Team



Yiran Li
Post-doc



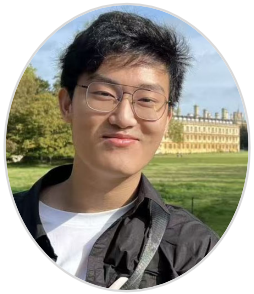
Yifang Tian
PhD Candidate



Mengli Duan
PhD Student



Ruiwu Liu
PhD Student



Yuanchen Gao
Visiting PhD Student
HKUST



Yifeng He
Master Student



Yingjian Bai
Master Student



Yaming Liu
Ex master Student
Now ByteDance

UNIVERSITY OF TORONTO

MIDDLEWARE SYSTEMS RESEARCH GROUP MSRG.ORG

AMA!

Sponsors:

Chaires de recherche du Canada / Canada Research Chairs

Canada

Ontario Research Fund

NSERC CRSNG



https://arxiv.org/abs/2508.12472

Bell

https://msrg.github.io/research/rca-aiops/

