TrioXpert: An Automated Incident Management Framework for Microservice System

Yongqian Sun[1], Yu Luo[1], Xidao Wen[2], Yuan Yuan[3],
Xiaohui Nie[4], Shenglin Zhang[1], Tong Liu[5], Xi Luo[5]

[1]Nankai University, [2]BizSeer,
[3]National University of Defense Technology,
[4]Computer Network Information Center, Chinese Academy of Sciences, [5]Lenovo

Outline

- Background & Motivation
- Framework Design
- Evaluation

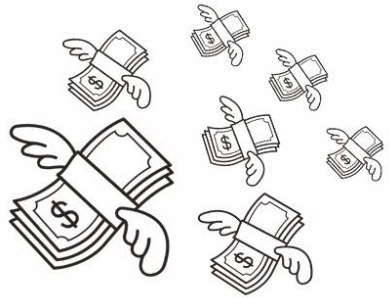
Microservices Systems

Microservice architectures have become the standard for modern enterprise systems

Impact of Incidents

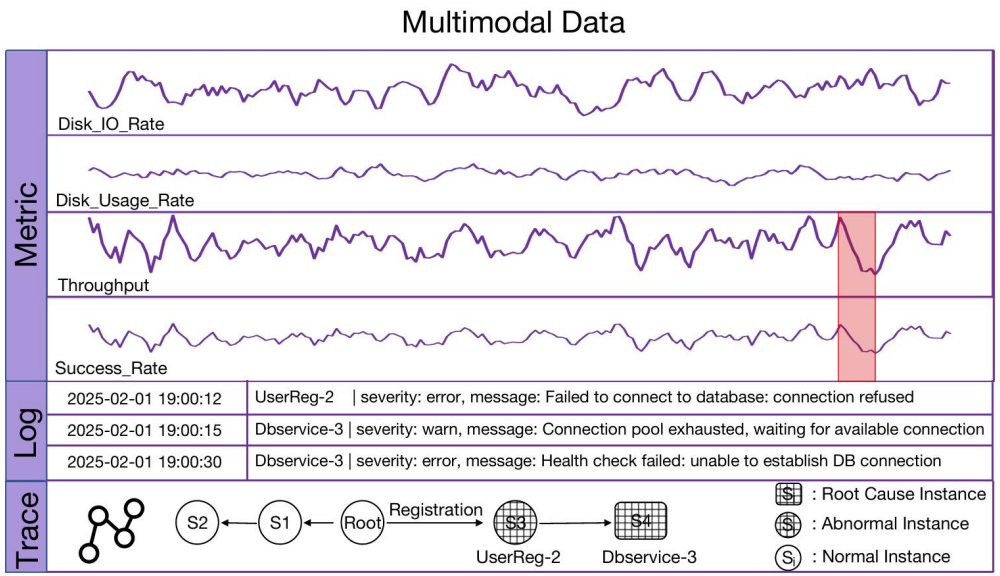
Incidents

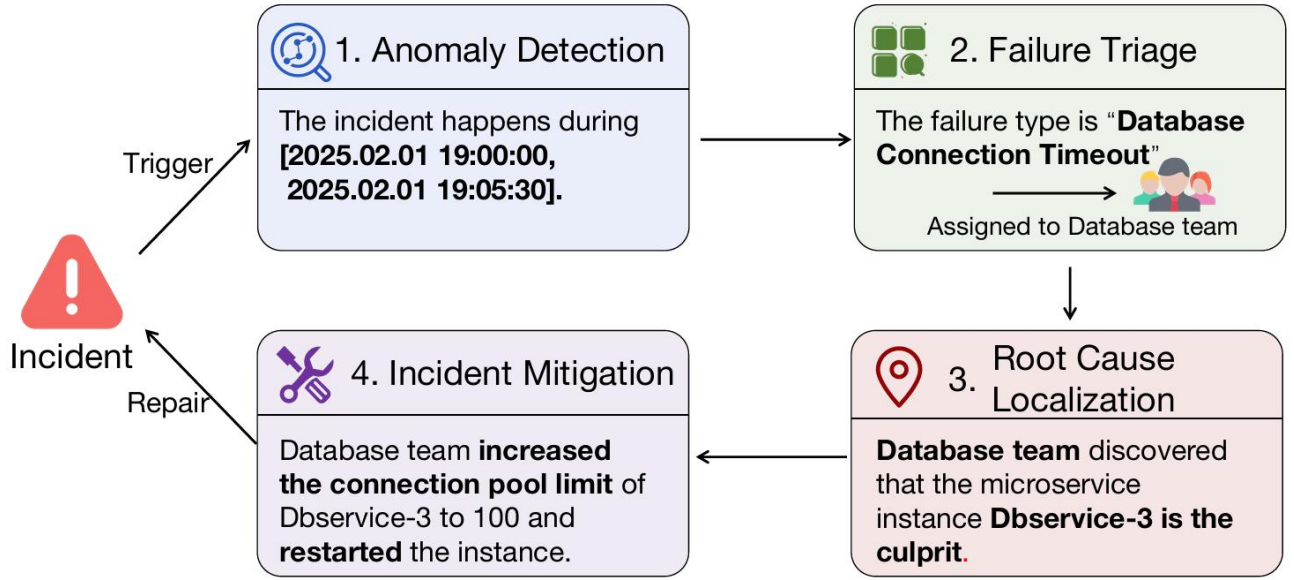
Poor UX



Financial Loss

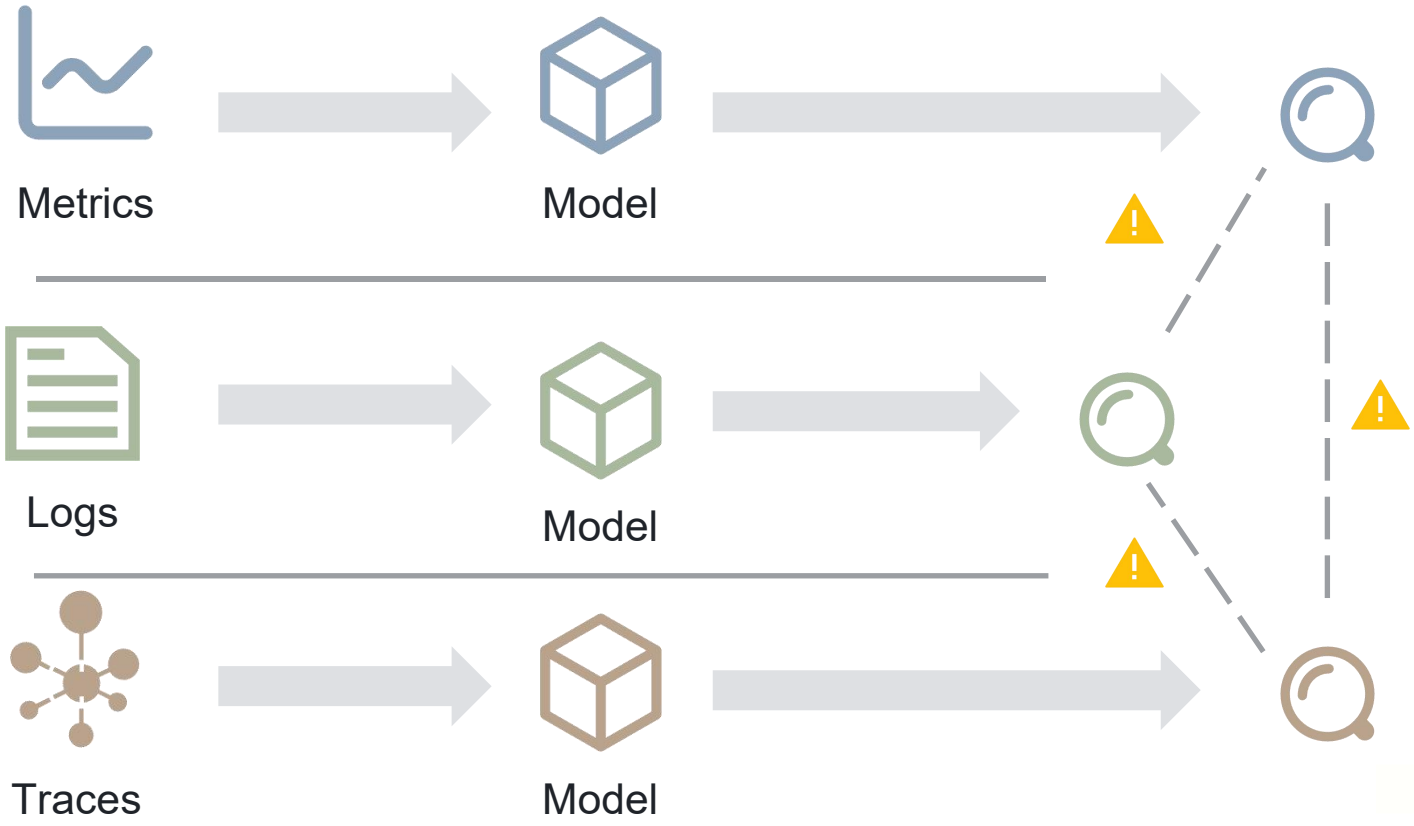
Typical Lifecycle of Incident Management





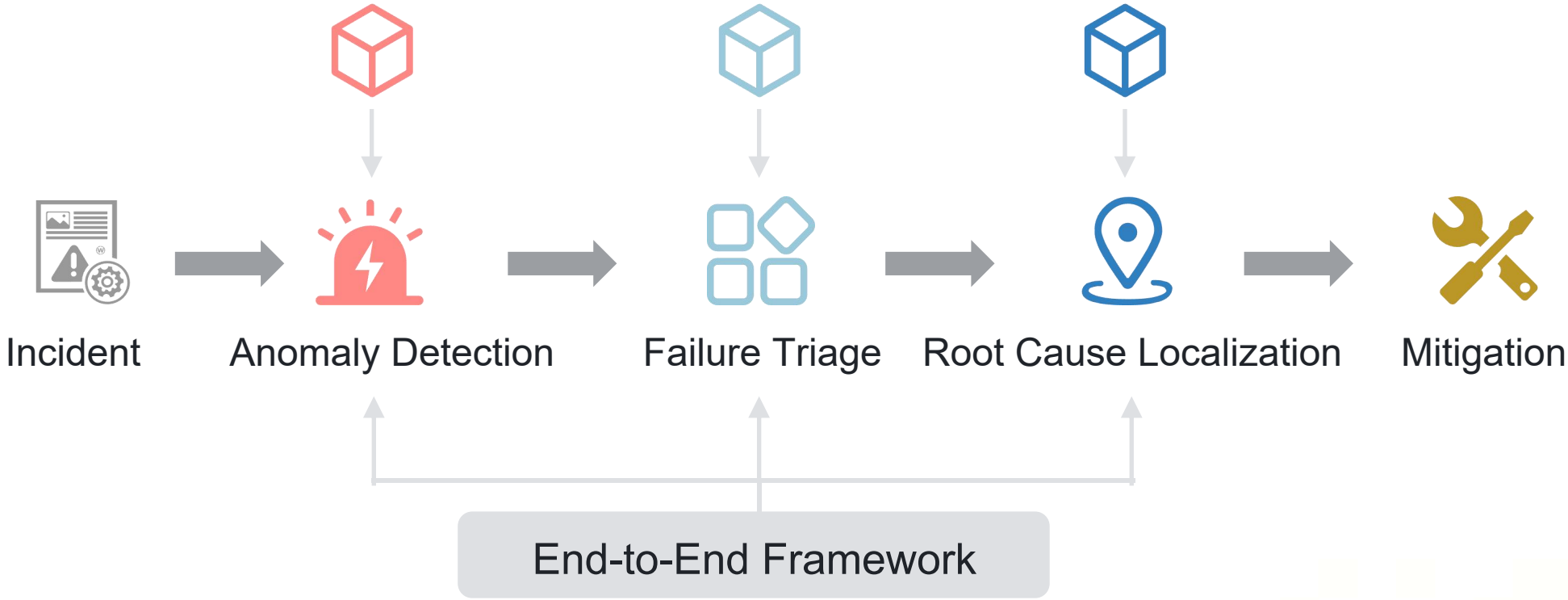
OCEs require automated management

Why We Need an End-to-End Framework?



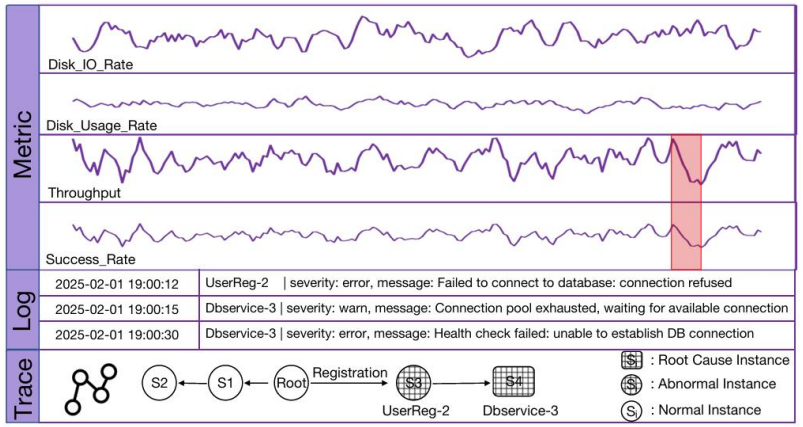
Inconsistent or partial conclusions

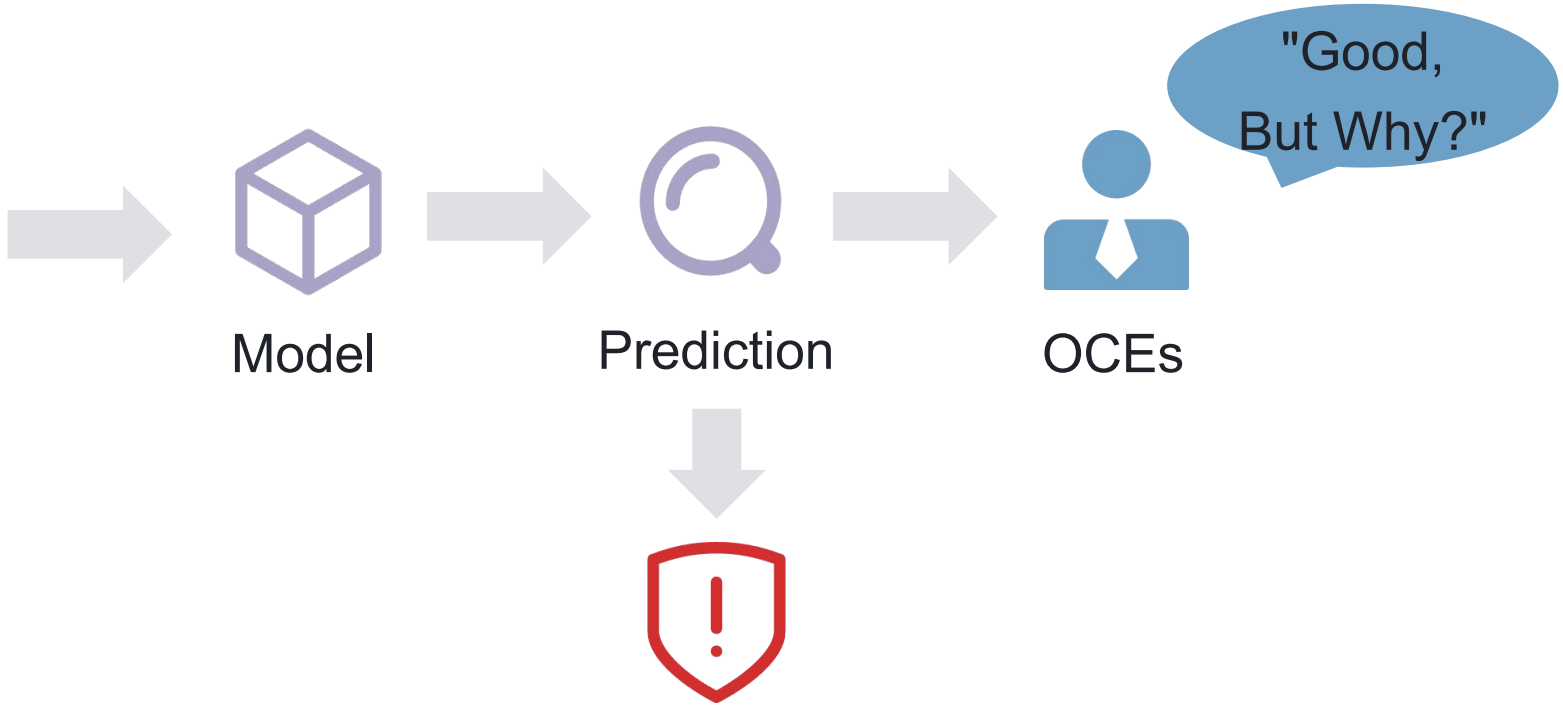
Why We Need an End-to-End Framework?



Increased Deployment Costs & Integration Overhead

Can We Trust a Black-Box in Critical Operations?





No Explanation, No Trust

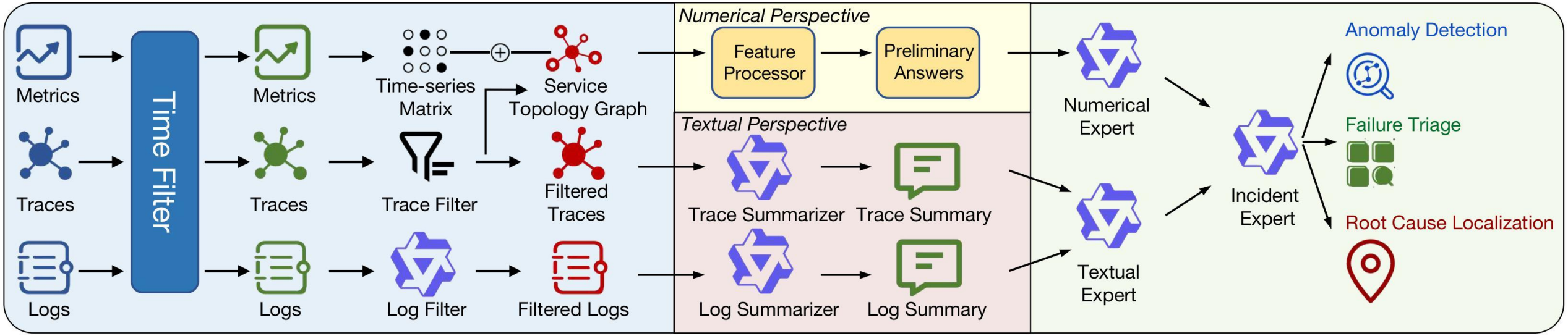
Challenges when integrating LLMs

- Semantic impoverishment in multimodal fusion
- Textual data overload in real-time incident management
- LLM limitations in complex and trust critical incident management

Outline

- Background & Motivation
- Framework Design
- Evaluation

TrioXpert Overview



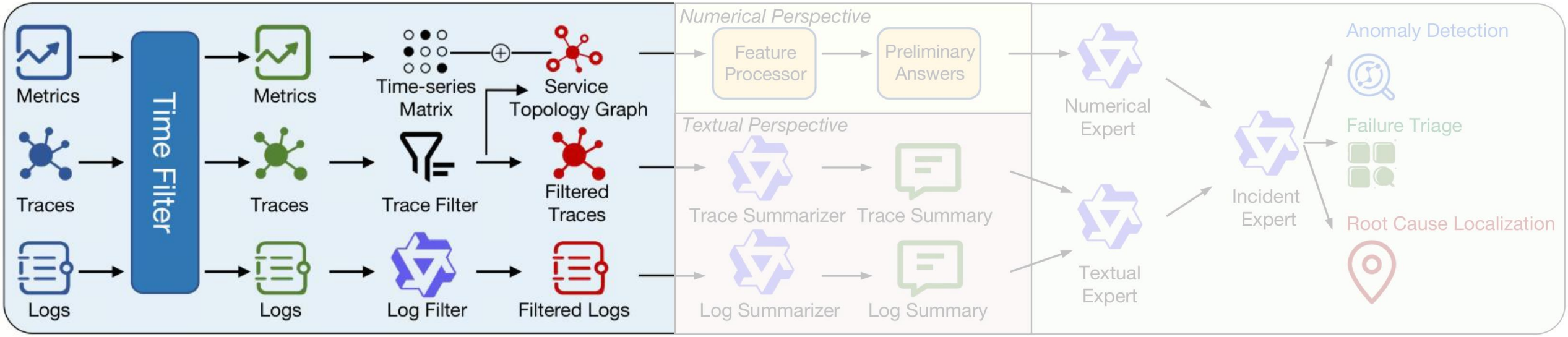
(a). Multimodal Data Preprocessing

(b). Multi-Dimensional System Status Representation

(c). LLMs Collaborative Reasoning

An End-to-End Incident Management Framework for Microservice System

TrioXpert - Module #1



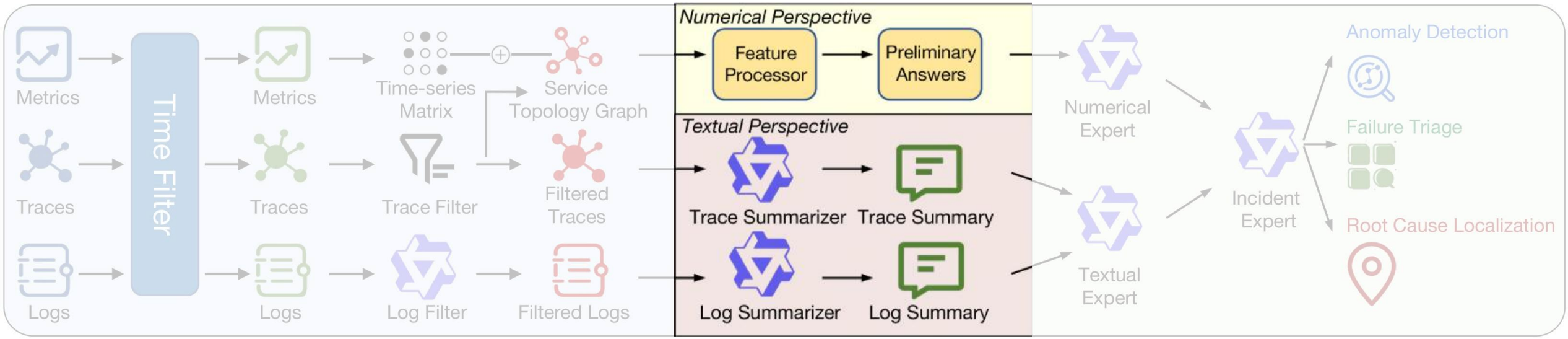
(a). Multimodal Data Preprocessing

(b). Multi-Dimensional System Status Representation

(c). LLMs Collaborative Reasoning

Module #1 Multimodal Data Preprocessing

TrioXpert - Module #2



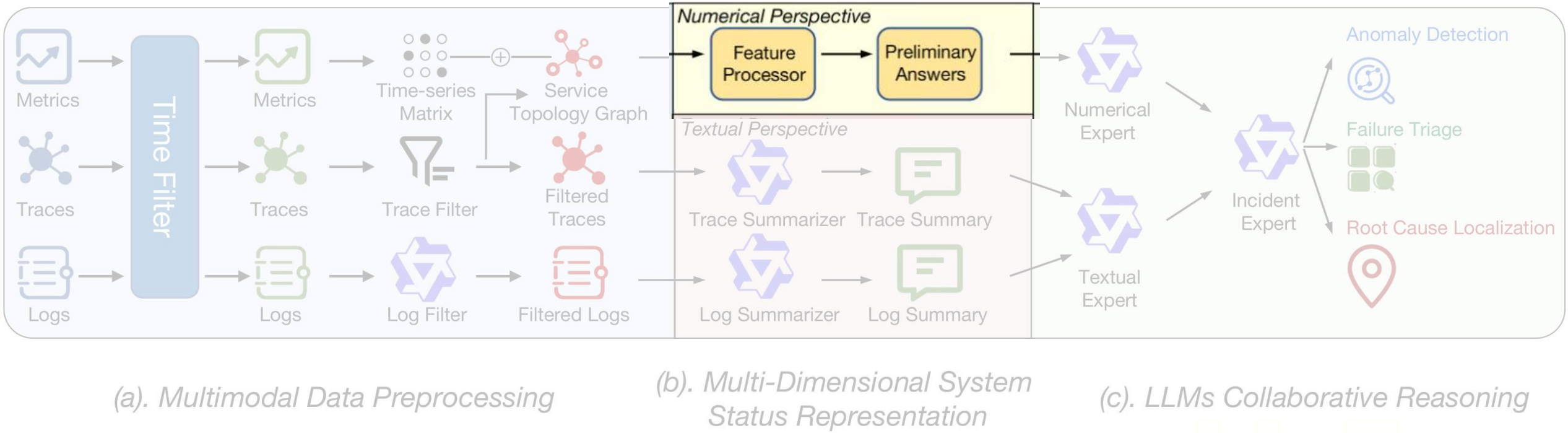
(a). Multimodal Data Preprocessing

(b). Multi-Dimensional System Status Representation

(c). LLMs Collaborative Reasoning

Module #2 Multi-Dimensional System Status Representation

TrioXpert - Module #2



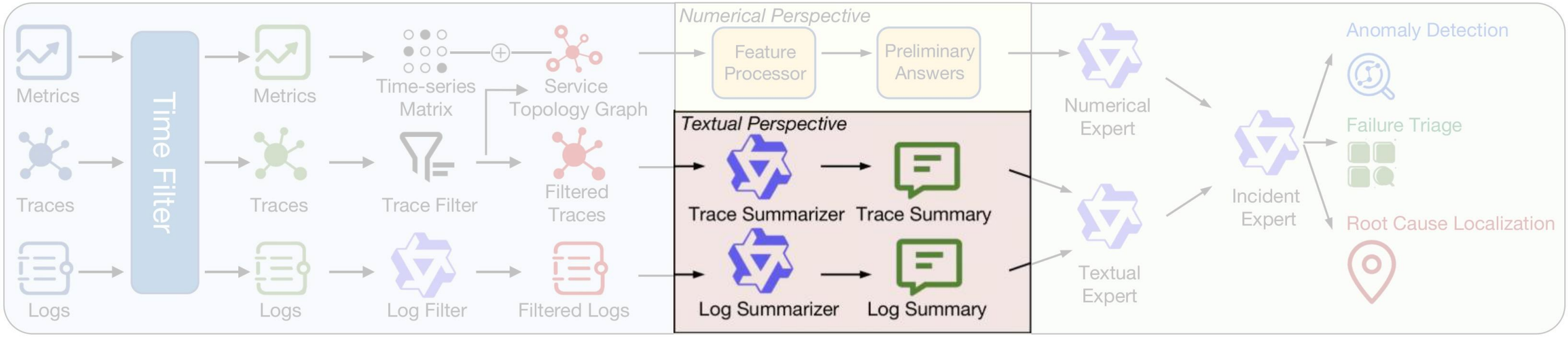
(a). Multimodal Data Preprocessing

(b). Multi-Dimensional System Status Representation

(c). LLMs Collaborative Reasoning

Module #2 Numerical Perspective

TrioXpert - Module #2



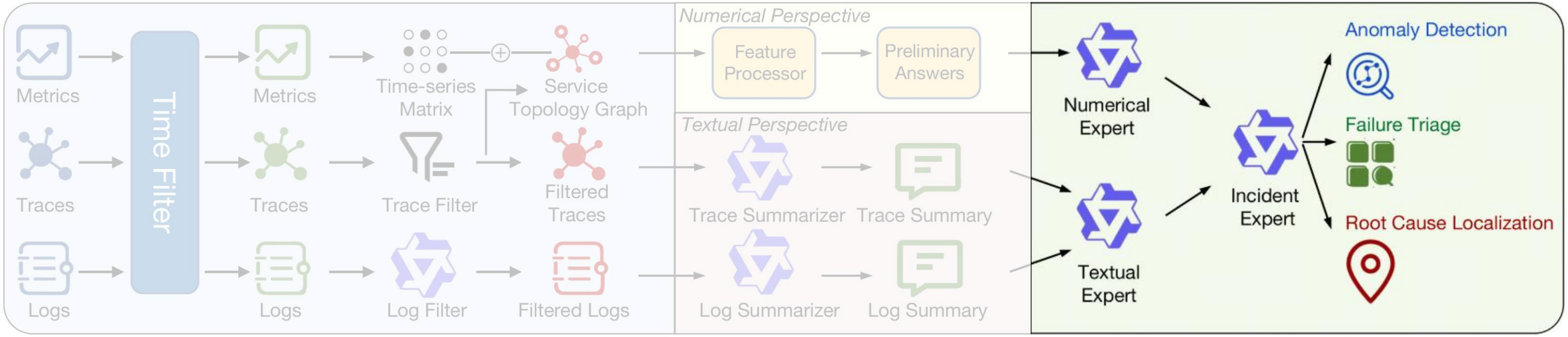
(a). Multimodal Data Preprocessing

(b). Multi-Dimensional System Status Representation

(c). LLMs Collaborative Reasoning

Module #2 Textual Perspective

TrioXpert - Module #3



(a). Multimodal Data Preprocessing

(b). Multi-Dimensional System Status Representation

(c). LLMs Collaborative Reasoning

Module #3 LLMs Collaborative Reasoning

Outline

- Background & Motivation
- Framework Design
- Evaluation

Evaluation: Performance

TABLE II
PERFORMANCE COMPARISON ON AD, FT, RCL, AND TIME. "-" MEANS THIS METHOD DOES NOT COVER THE TASK.

Methods	$\mathscr{D}1$										$\mathscr{D}2$									
	AD			FT			RCL			Efficiency	AD			FT			RCL			Efficiency
	Precision	Recall	F1	Precision	Recall	F1	Top@1	Top@3	Avg@5	Time (s)	Precision	Recall	F1	Precision	Recall	F1	Top@1	Top@3	Avg@5	Time (s)
TrioXpert	0.880	0.972	**0.924**	0.852	0.768	**0.807**	0.651	0.778	**0.773**	14.314	0.854	0.972	**0.909**	0.814	0.725	**0.767**	0.550	0.775	**0.750**	12.597
ART [1]	0.759	0.621	0.683	0.786	0.794	0.790	0.683	0.762	0.757	0.872	0.593	0.972	0.737	0.860	0.650	0.740	0.375	0.825	0.738	1.363
DiagFusion [2]	-	-	-	0.675	0.500	0.574	0.310	0.452	0.467	4.145	-	-	-	0.797	0.527	0.634	0.582	0.709	0.695	3.297
Eadro [4]	0.425	0.946	0.586	-	-	-	0.137	0.315	0.302	0.627	0.767	0.935	0.842	-	-	-	0.157	0.315	0.310	0.899
Hades [29]	0.866	0.863	0.865	-	-	-	-	-	-	0.104	0.867	0.868	0.868	-	-	-	-	-	-	0.415
MicroCBR [11]	-	-	-	0.667	0.796	0.726	-	-	-	0.278	-	-	-	0.629	0.678	0.653	-	-	-	0.306
PDiagnose [30]	-	-	-	-	-	-	0.615	0.692	0.685	4.342	-	-	-	-	-	-	0.037	0.296	0.285	9.919

Evaluation: Ablation Study

A1	Remove the textual pipelines
A2	Remove the numerical pipeline
A3	Replace the multi-expert reasoning with a single LLM
A4	Disable conflict resolution and aggregation
A5	Disable hallucination mitigation

Evaluation: Ablation Study

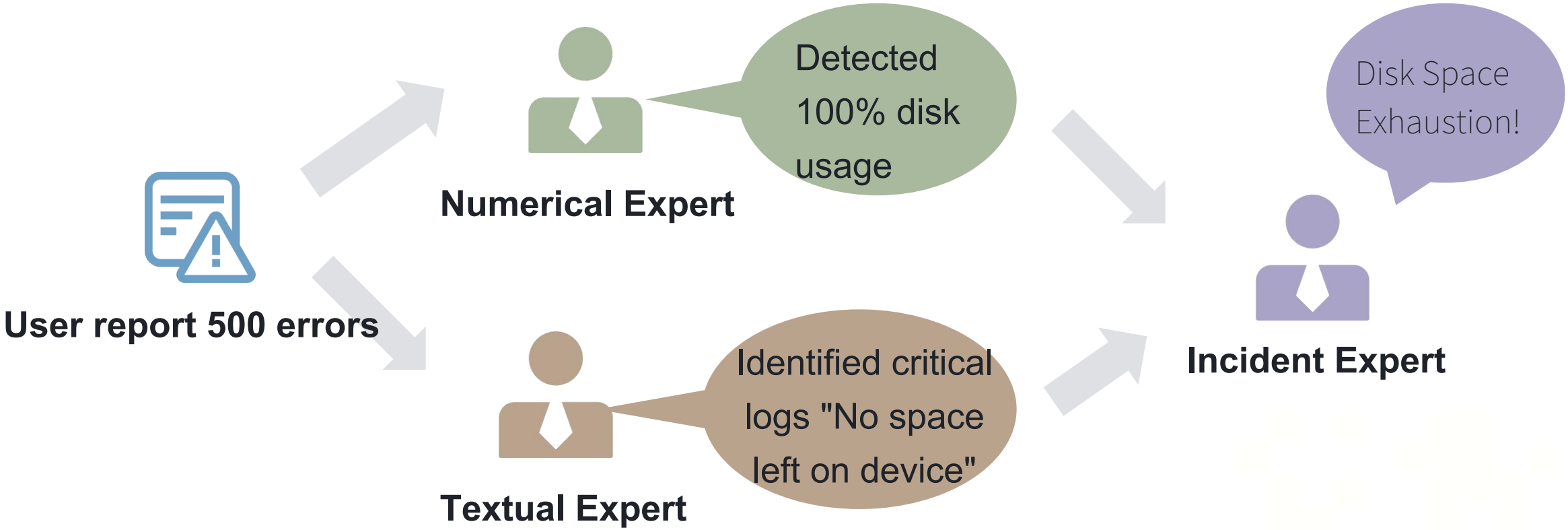
TABLE III
THE EVALUATION RESULTS OF ABLATION STUDY.

Methods	$\mathscr{D}1$			$\mathscr{D}2$		
	AD: F1	FT: F1	RCL: Avg@5	AD: F1	FT: F1	RCL: Avg@5
TrioXpert	**0.924**	**0.807**	**0.773**	**0.909**	**0.767**	**0.750**
$\mathcal{A}1$	0.725	0.190	0.667	0.832	0.685	0.625
$\mathcal{A}2$	nan	0.261	0.238	nan	0.352	0.275
$\mathcal{A}3$	0.672	0.398	0.534	0.583	0.284	0.608
$\mathcal{A}4$	0.428	0.294	0.397	0.552	0.359	0.517
$\mathcal{A}5$	0.339	0.157	0.362	0.405	0.287	0.233

Case Study: Real-World Incidents from Lenovo Production

Traditional Method	TrioXpert
3 OCEs	Automated
2.5 h	26 s
5+ attempts	2 attempts
Manual reasoning	Interpretable reasoning chain

Case Study: Real-World Incidents from Lenovo Production



Lenovo OCEs Validation: Root cause was transparent and traceable

Thank you!