Improving the Reliability of Software-Intensive Infrastructures Using AIOps

Prof. Wahab Hamou-Lhadj

Concordia University

Montreal, Canada

wahab.hamou-lhadj@concordia.ca

LTB@ICPE'224

London, UK, 2024

AI for IT Operations

- AIOps relies on data analytics and machine learning to automate and optimize IT operations.

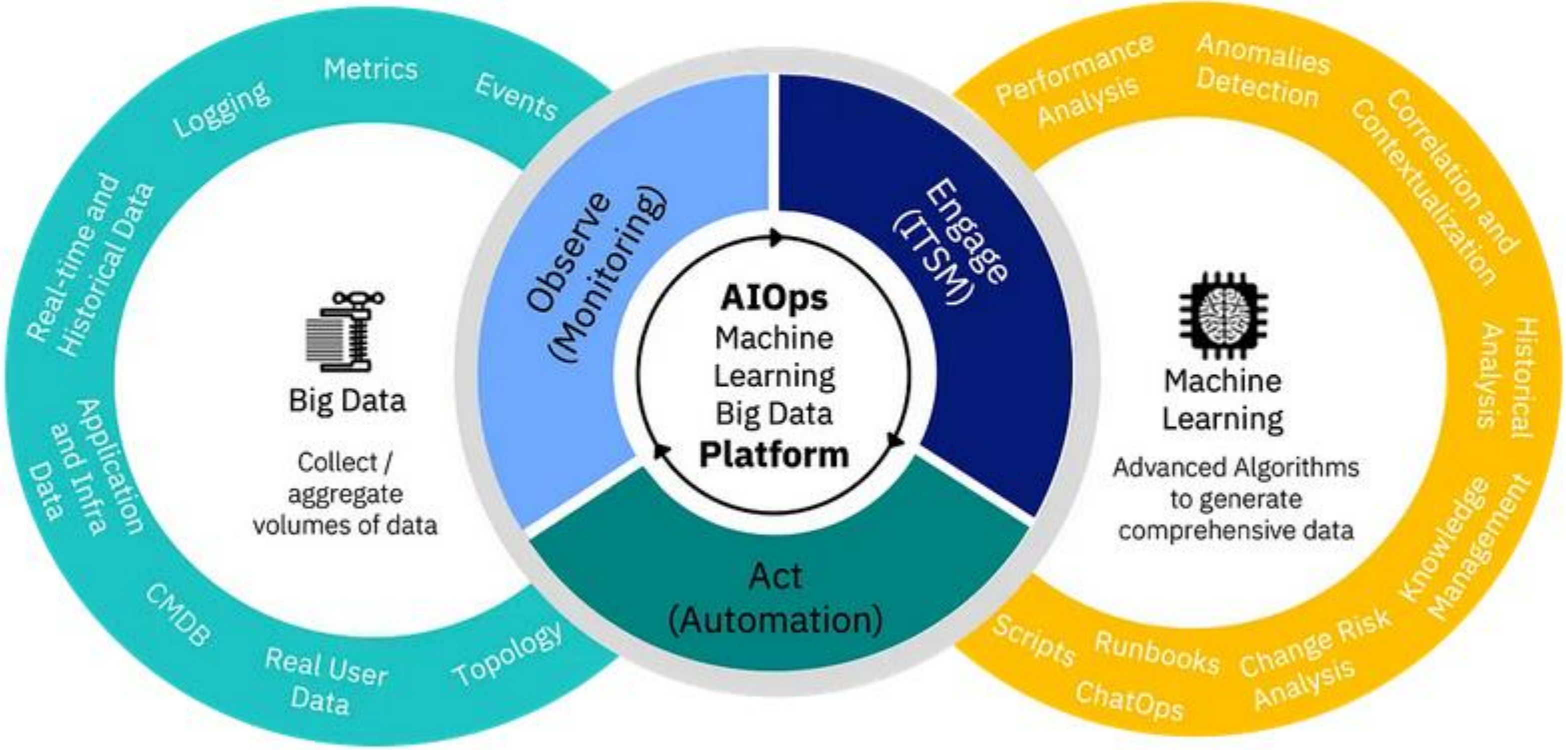
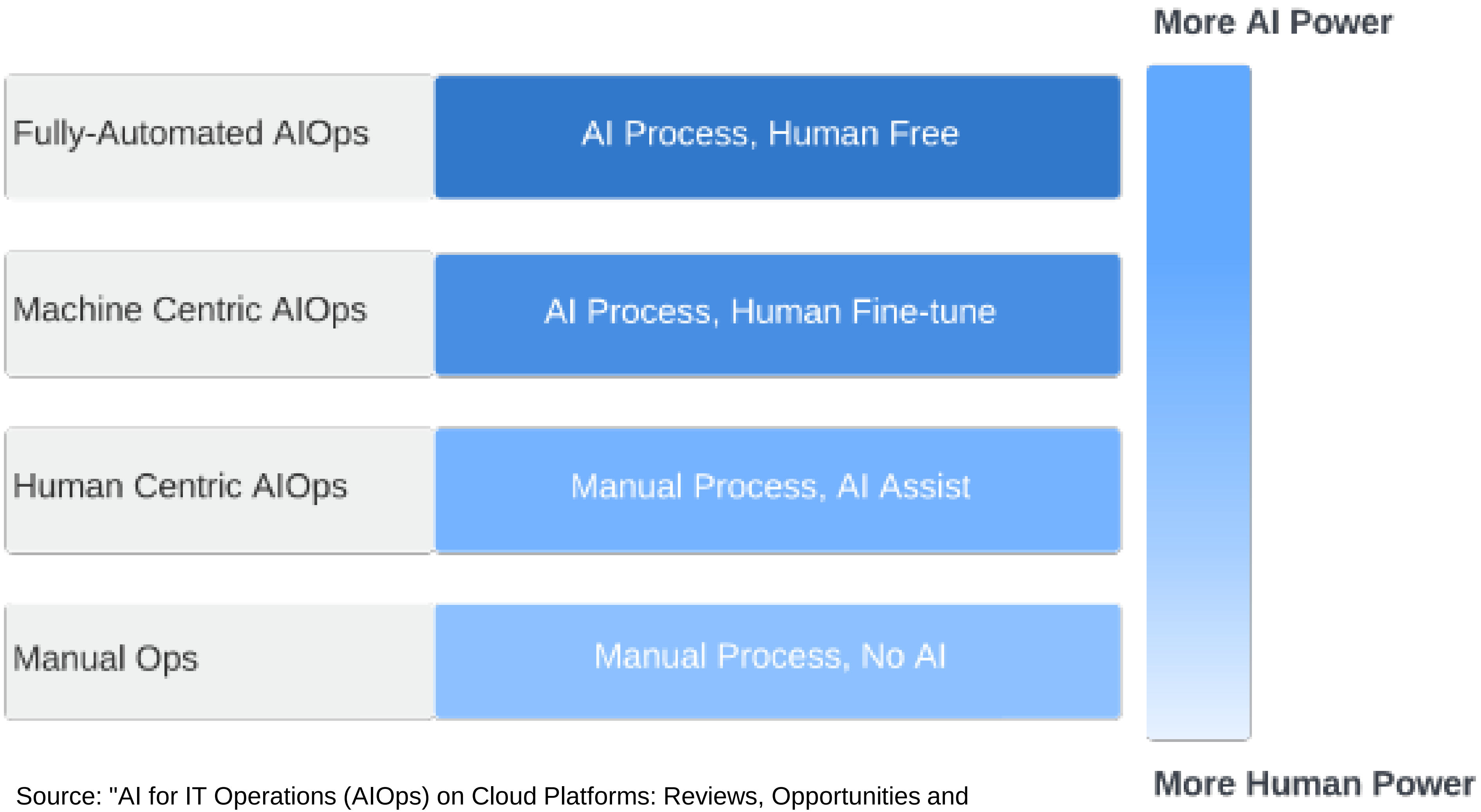


Figure source: Achieving AIOps with Instana, by Tiago Dias Generoso URL: https://tiagodiasgeneroso.medium.com/achieving-aiops-with-instana-1453a6dc5456

AIOps Transformation



Source: "AI for IT Operations (AIOps) on Cloud Platforms: Reviews, Opportunities and Challenges," by Cheng et al.

Why we need AIOps?

- Shift towards **DevOps and CI practices**
- **Operational complexity** of IT infrastructures
- Emergence of advanced **distributed architectures**
- Increasing **reliance on IT operations** tools
- Emergence of **new AI paradigms** (e.g., LLMs)
- Challenges **hiring and retaining workforce**

Software Observability

- **In control theory:**
  - **Observability** is “a measure of how well internal states of a system can be inferred from knowledge of its external outputs” [Wikipedia]

- **Software Observability:**
  - A set of end-to-end techniques and processes that allow us **to reason** about **what a software system is doing and why** by analyzing its external outputs.

Monitoring vs Observability

- **Monitoring:**
  - Tracks known metrics and raises alerts when thresholds are not met
  - Four golden signals of Google SRE: latency, traffic, errors, and saturation
  - Answers the question: “how is the system doing?”
  - Helps diagnose known problems

- **Observability:**
  - Answers the question: “what is the system doing and why?”
  - Enables to reason about the system by observing its outputs
  - Helps diagnose known and unknown problems

Monitoring vs Observability

- **Monitoring:**
  - Tracks known metrics and raises alerts when thresholds are not met
  - Four golden signals of Google SRE: latency, traffic, errors, and saturation
  - Answers the question: “how is the system doing?”
  - Helps diagnose known problems

- **Observability:**
  - Answers the question: “what is the system doing and why?”
  - Enables to reason about the system by observing its outputs
  - Helps diagnose known and unknown problems

A 2022 study by AppDynamics shows that 91% of participants believe that gaining full observability into their systems would be revolutionary for their business

A VMware report shows that traditional monitoring tools are not enough to understand today’s complexity of large-scale systems

Focus of Current Research

Logging, Tracing, and
Date Management

Anomaly
Detection

IR Management, RCA,
Mitigation techniques

Data Privacy &
Regulatory Compliance

The Log Parsing Problem

- Logs are largely unstructured
- Automatic extraction of log templates is a complex problem because:
  - A typical file may contain thousands of log templates
  - Systems contain many types of log data
  - Lack of logging guidelines and standards

Logging Statement: `LOG.info("Received Block "+ block_id + " of size " + block_size + " from " + ip)`

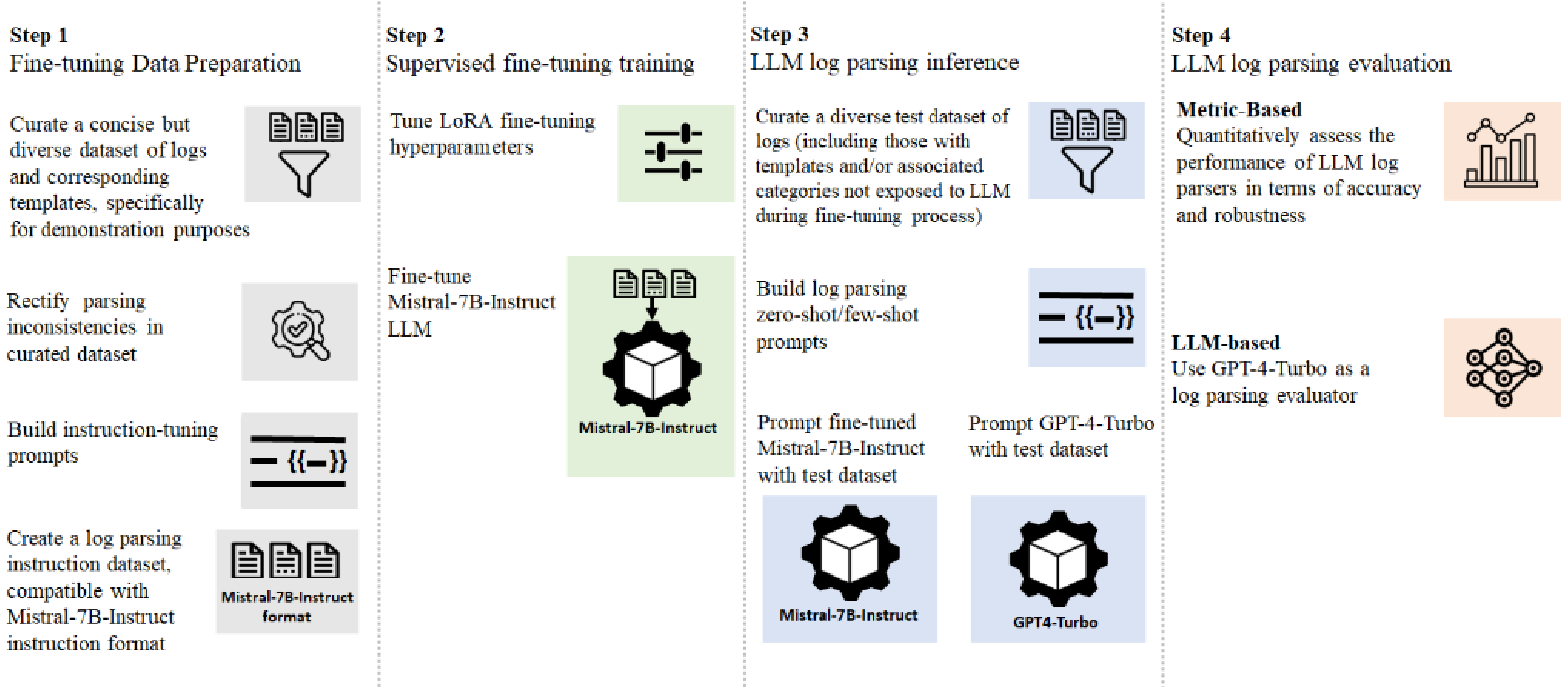
Log Event: `270423 283349 9876 INFO dfs.DataNodeResponder: Received block blk_-1680 of size 4536 from 10.163.23.167`

Log Template: `Received Block <*> of size <*> from <*>`

Log Parsing with LLMs

- LLMs have been used for automatic generation of logging statements, log parsing, and root cause analysis
- LLM-based log parsing studies have mainly leveraged general-purpose LLMs such as ChatGPT
- The use of such LLMs for log analytics pose three challenges:
  - Privacy: Using a proprietary LLM (e.g., GPT-4) increases the risk of violating privacy regulations
  - Tool Integration: Integrating a third-party LLM with existing log analytics tooling can be challenging
  - Cost: The high-performing LLMs tend to be expensive when used with large data

LLM-based Log Parsing Approach Using Mistral-7B



Findings

- RQ1: What is the accuracy of Mistral-7B compared to GPT-4 across various configuration settings using a metric-based evaluation method?

Findings

- RQ1: What is the accuracy of Mistral-7B compared to GPT-4 across various configuration settings using a metric-based evaluation method?

Model	MLA	ED		F1 Score	
		Mean	Median	Mean	Median
Mistral-7B (0-shot)	22.9%	21.9	12.0	0.23	0.0
Mistral-7B (2-shot)	14.1%	54.2	55.5	0.13	0.0
Mistral-7B (Fine-tuned)	74.8%	7.2	0.0	0.74	1.0
GPT-4 (0-shot)	47.2%	6.4	2.0	0.46	0.0
GPT-4 (2-shot)	72.2%	9.2	0.0	0.71	1.0

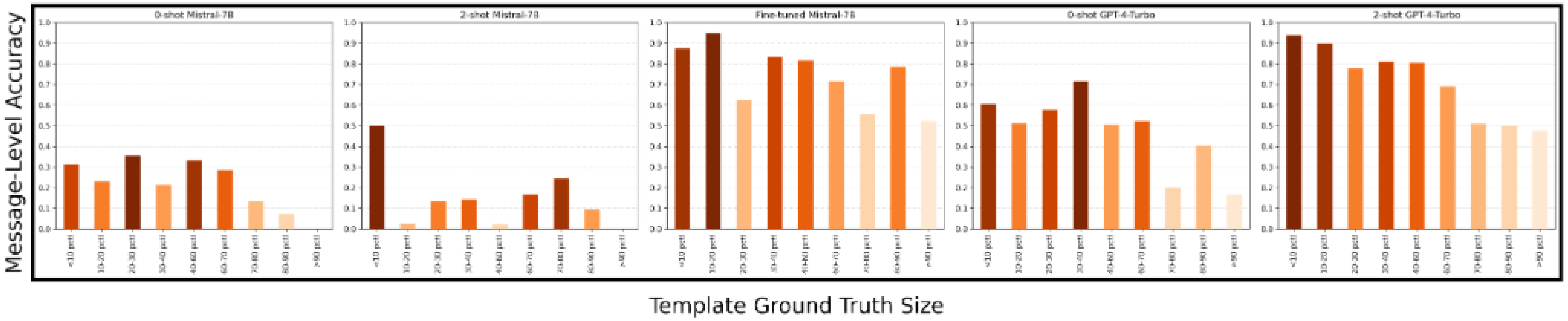
The results show that fine-tuned Mistral-7B achieves better accuracy compared to GPT-4 using all three metric-based assessment, MLA, ED, and F1 Score

Findings

- RQ2: What is the robustness of Mistral-7B compared to GPT-4 across various configuration settings using a metric-based evaluation method?

Findings

- RQ2: What is the robustness of Mistral-7B compared to GPT-4 across various configuration settings using a metric-based evaluation method?



- The results show that fine-tuned Mistral-7B achieves the best robustness in metric-based assessment with different template sizes and different datasets.
- It also has a satisfactory robustness when used with familiar datasets. However, it requires enhancement in order to be more robust when used with new and unseen log files.

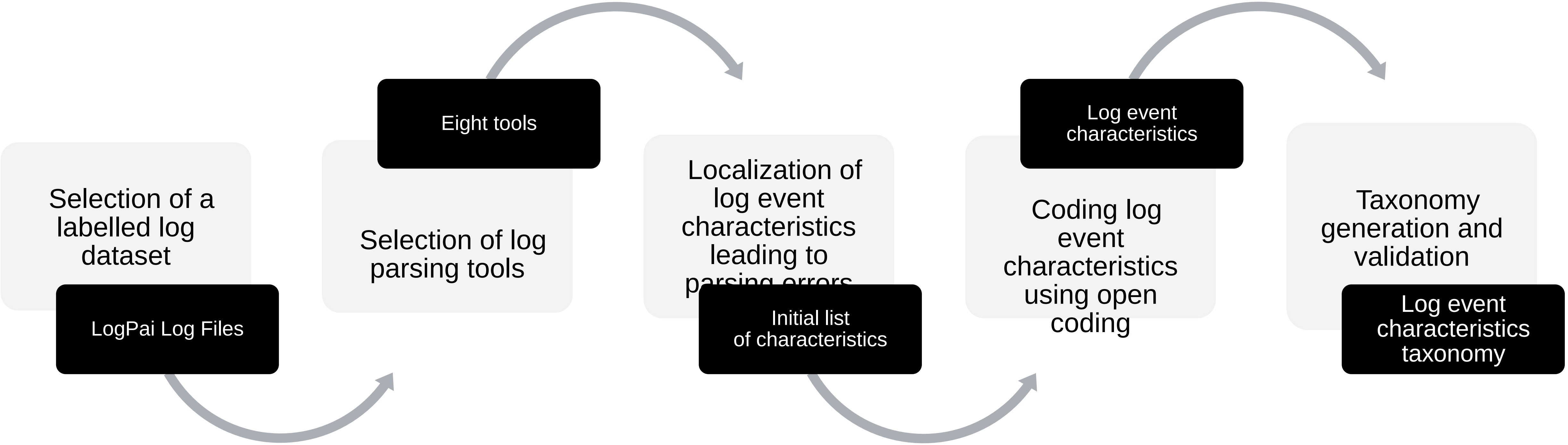
Findings

- RQ3: What is the accuracy and robustness of Mistral-7B compared to GPT-4 using an LLM-based evaluation method?

Findings

- RQ3: What is the accuracy and robustness of Mistral-7B compared to GPT-4 using an LLM-based evaluation method?
  - We found that fine-tuned Mistral-7B achieves the best robustness in metric-based assessment with different template sizes and different datasets
  - It has a satisfactory robustness when used with familiar datasets
  - It requires enhancement in order to be more robust when used with new and unseen log files

A Taxonomy of Log Parsing Errors



Log Event Characteristics

Log Event Characteristics grouped into categories

Data Types Category	Structural Patterns Category	Log Message Composition Category
(1) Datetime tokens; (2) Time duration tokens; (3) Decimal; (4) Data Volume and Unit; (5) Protocol name; (6) MAC Address; (7) Non-standard MAC Address Format; (8) ID Token; (9) Boolean; (10) IPv4 token; (11) IPv6 token; (12) Domain name; (13) Use of Nouns; (14) Hexadecimal; (15) URL; (16) Boolean in a format other than True/False; (17) URL with Query parameters; (18) Folder Structure; (19) Use of UUID.	(20) Single-level Nested Tokens; (21) Multi-level Nested Tokens; (22) Equals-separated Key-Value Pairs; (23) Colon-Delimited Key-Value Pairs; (24) Word-number pair; (25) Enclosed Quotations; (26) Unseparated Token Sequence.	(27) Alphanumeric and Special Characters; (28) Token with Punctuation Marks; (29) Log event with only static tokens; (30) Log highlighters.

LECs with highest impact on log parsing tools

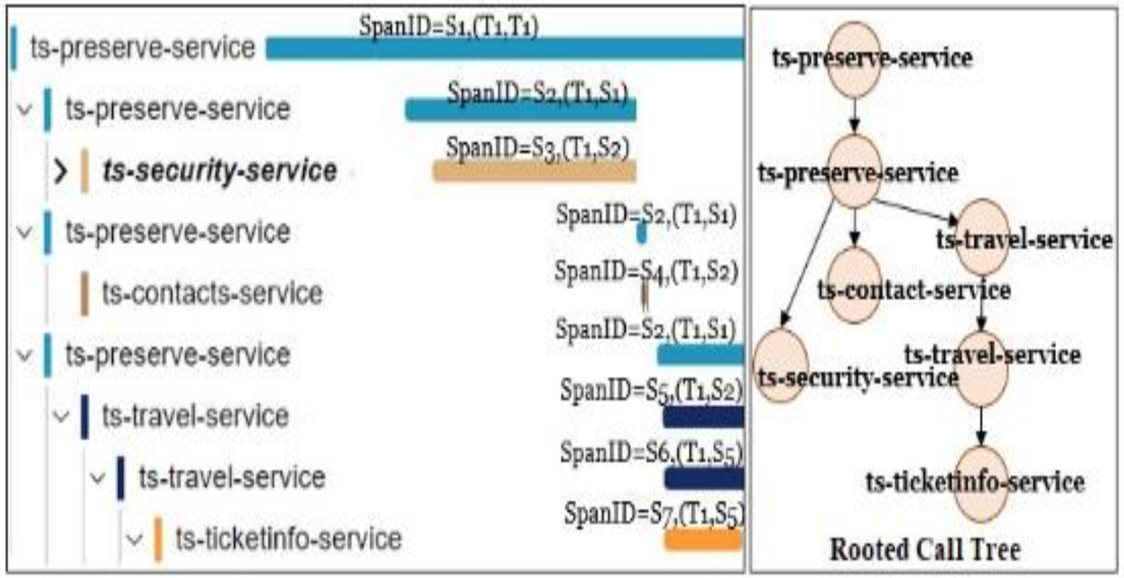
Category	LEC	AEL	Drain	Iplom	Lenma	Logmine	Shiso	Spell	ULP
Data types	Decimal	8.06%	7.61%	8.45%	7.43%	8.29%	7.79%	7.56%	8.55%
	IPv4 token	6.01%	5.43%	6.50%	5.40%	5.76%	5.84%	5.49%	6.51%
	Datetime tokens	5.16%	4.87%	5.29%	5.65%	5.64%	5.11%	5.18%	4.98%
Structural patterns	Unseparated Token Sequence	19.83%	19.44%	19.54%	19.76%	19.58%	20.02%	19.52%	19.08%
	Key-value pairs with a colon	11.28%	11.42%	10.99%	10.68%	10.82%	10.63%	11.41%	11.05%
Log message composition	Alphanumeric & Special Characters	12.28%	13.02%	11.87%	12.28%	11.02%	11.95%	12.17%	12.73%

ServiceAnomaly: Anomaly Detection in Microservices Using Distributed Traces and Profiling Metrics

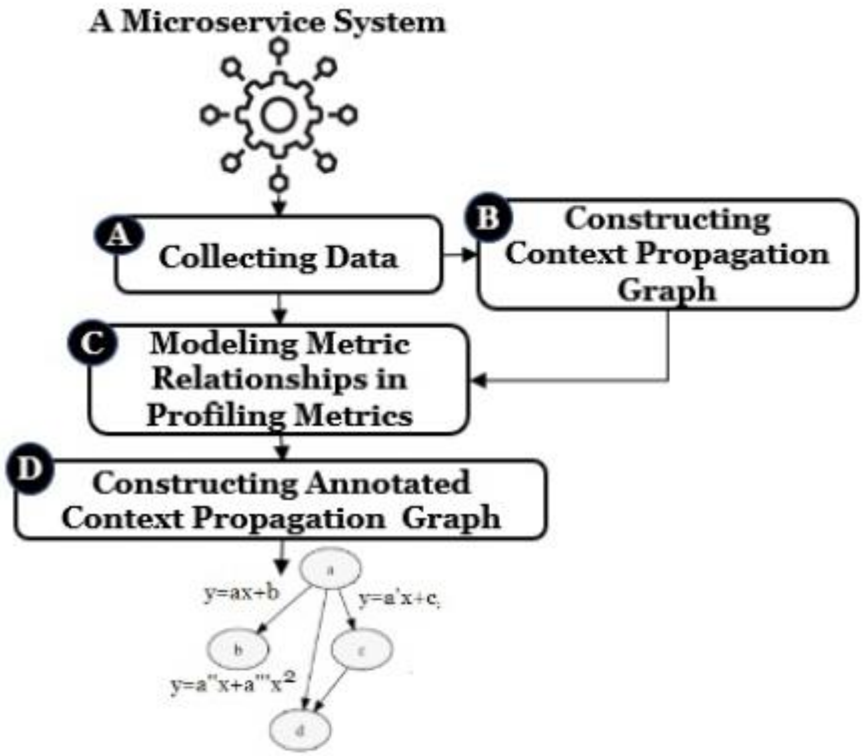
- A distributed trace represents an end-to-end request and contains a series of events generated from a microservice-based systems

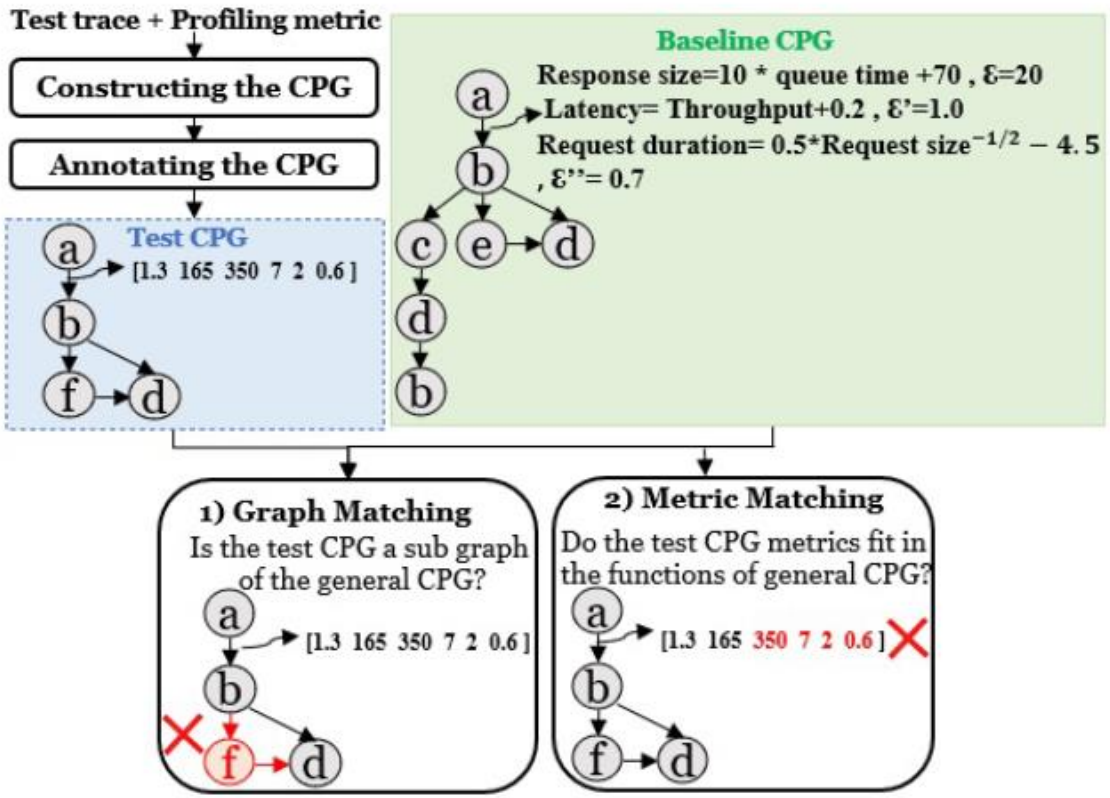
- ServiceAnomaly combines service dependency graphs with multiple metrics to create a context propagation graph

- It helps detect and analyze the causes of anomalies.



ServiceAnomaly Approach





Evaluation

- RQ1. How accurate is ServiceAnomaly at detecting anomalies?

Evaluation

- RQ1. How accurate is ServiceAnomaly at detecting anomalies?

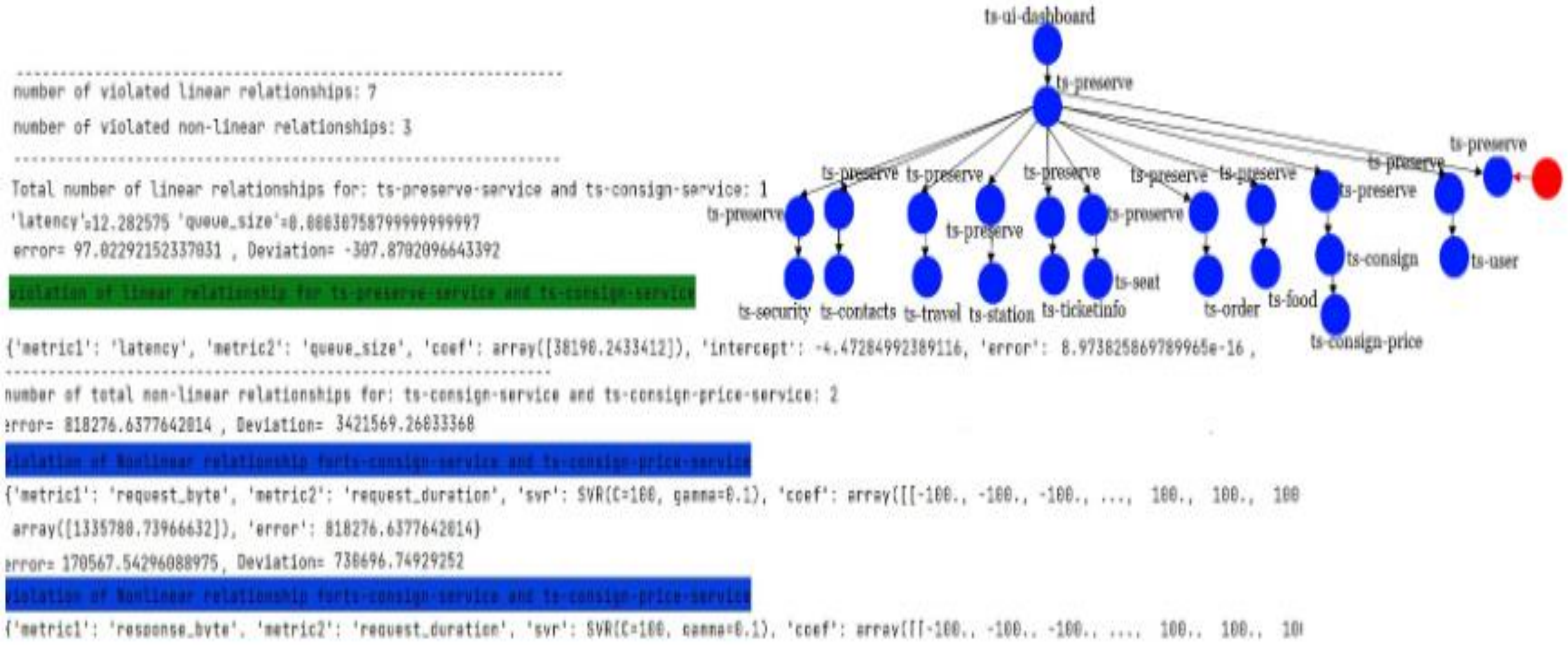
  - We found that ServiceAnomaly can detect anomalies with an F1-score up to 85% for TeaStore and 86% for TrainTicket
  - The RMSE error evaluation metric yields a more accurate model for both systems compared to other error evaluation metrics
  - We also showed that the combination of CPG and profiling metrics is an effective way to detect different types of faults

Evaluation

- RQ2. How can the ServiceAnomaly approach be used to analyze anomalies?

Evaluation

- RQ2. How can the ServiceAnomaly approach be used to analyze anomalies?



Anomaly Detection Techniques for AIOps

OmniAnomaly	A variational autoencoder framework enhanced by RNN
CAE-ensemble	Convolutional Autoencoder (CAE)
InterFusion	Hierarchical variational autoencoder (HVAE) architecture
MAD-GAN	Generator- discriminator architecture using LSTM-RNNs
SGmVRNN	A variational RNN (VRNN)
USAD	Encoder-decoder architecture trained within an adversarial training framework
ALAD	Bi-directional Generative Adversarial Networks (GANs)
BiLSTM	A bidirectional LSTM (BiLSTM)

Root Cause Analysis and Mitigation Using LLMs

- A study on “Recommending Root-Cause and Mitigation Steps for Cloud Incidents using Large Language Models” by Ahmed et al.
- A large-scale study in Microsoft on over 40,000 incidents from 1000+ cloud services with six semantic and lexical metrics.
- Fine-tuning significantly improves the effectiveness of LLMs for incident data.
- GPT-3.x significantly outperform encoder-decoder models in our experiments
- Manual inspection and validation with experts is needed to assess the actual performance

The Growing Field of Root Cause Analysis with LLMs

Exploring LLM-based Agents for Root Cause Analysis

Devjeet Roy, Xuchao Zhang, Rashi Bhave, Chetan Bansal, Pedro Las-Casas, Rodrigo Fonseca, Saravan Rajmohan

Automated Root Causing of Cloud Incidents using In-Context Learning with GPT-4

Xuchao Zhang, Supriyo Ghosh, Chetan Bansal, Rujia Wang, Minghua Ma, Yu Kang, Saravan Rajmohan

LLM-Enhanced Causal Discovery in Temporal Domain from Interventional Data

Peiwen Li, Xin Wang, Zeyang Zhang, Yuan Meng, Fang Shen, Yue Li, Jialong Wang, Yang Li, Wenweu Zhu

PACE-LM: Prompting and Augmentation for Calibrated Confidence Estimation with GPT-4 in Cloud Incident Root Cause Analysis

Dylan Zhang, Xuchao Zhang, Chetan Bansal, Pedro Las-Casas, Rodrigo Fonseca, Saravan Rajmohan

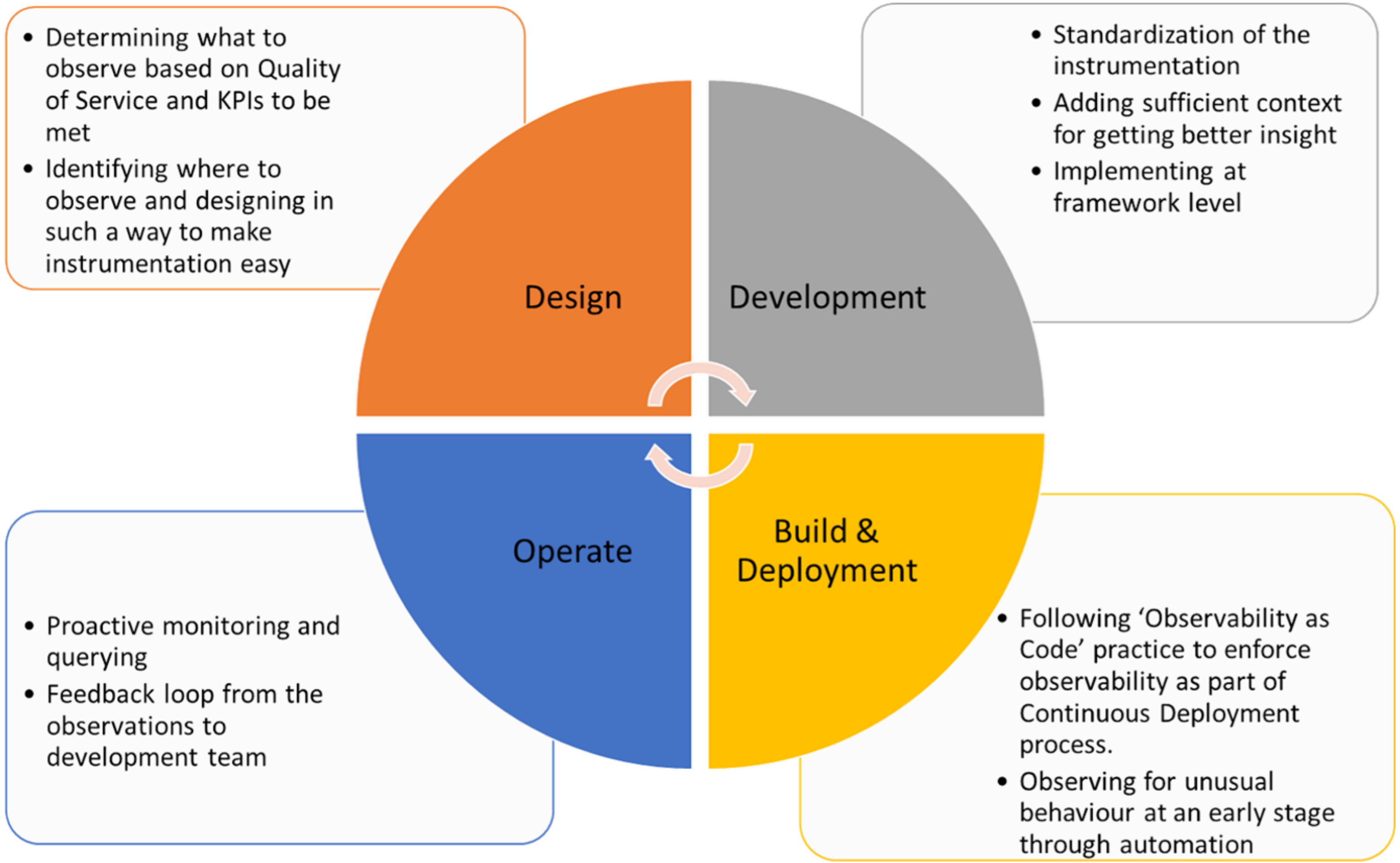
Challenges of using an AIOps solution in an organization

- **No standardized approach for AIOps**, limiting reuse and innovation
- Challenges **working with telemetry data** (size, structure, velocity, etc.)
- AIOps tools do not leverage **the full scale AI algorithms**
- **Lack of well established quality criteria** to assess the maturity of an AIOps solution
- Lack of **benchmark data** to compare solutions
- **Cost vs. benefit** is not well understood
- No clear alignment of **AIOps with a company's strategic directions**
- **Issues of governance, risk, and compliance** associated with AIOps
- **Roles and responsibilities** of AIOps operators are not well defined

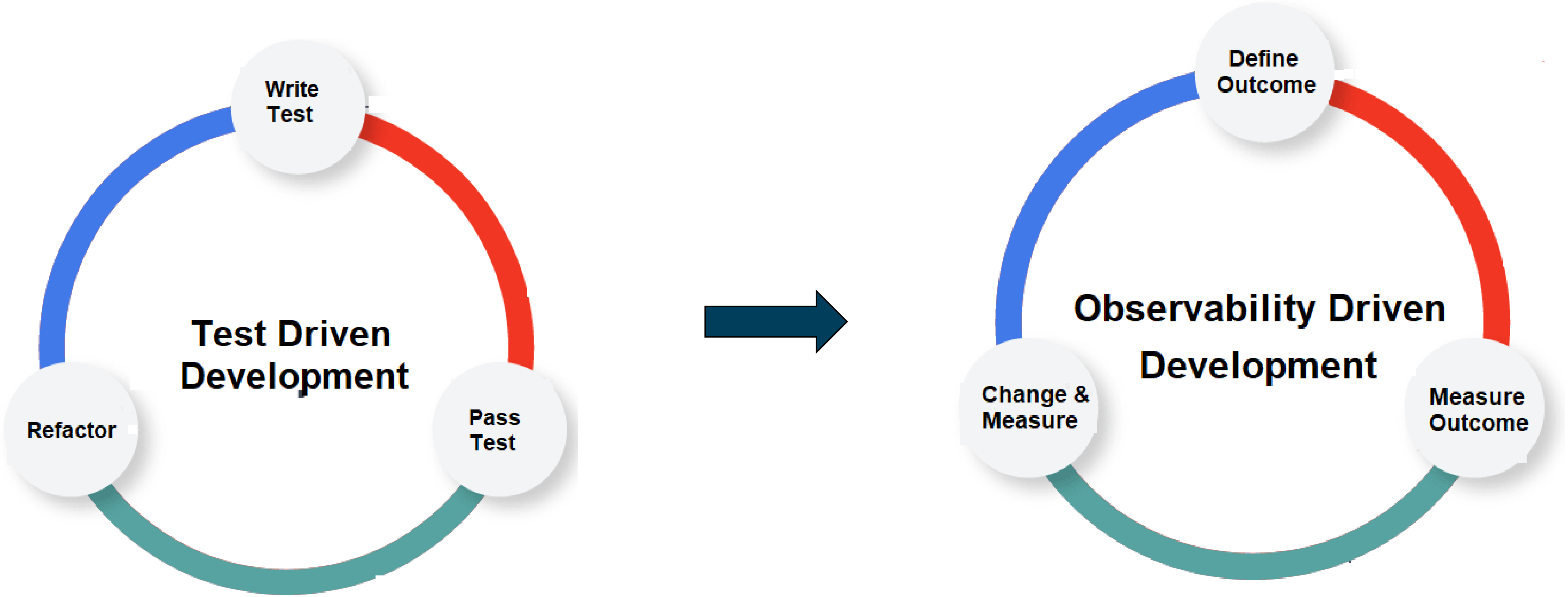
Observability-Driven Development

- Bringing observability **to early stages** of the software development lifecycle
- Defining a set of **observability patterns, best practices, and reusable solutions** to be used as guiding principles for developers
- A **systematic approach** to tracing, logging and profiling of software systems that considers different phases of the software process

Observability-Driven Development (cont.)

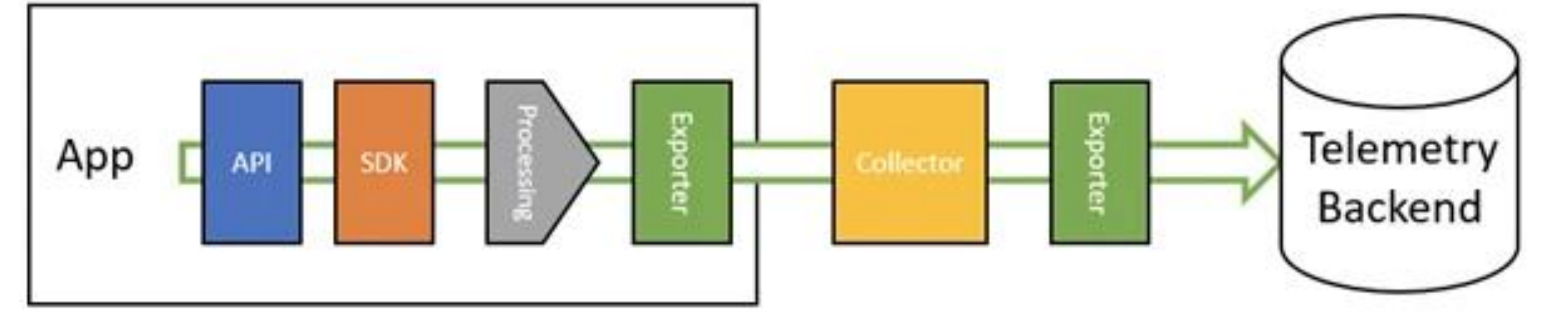


Observability-Driven Development (cont.)



OpenTelemetry Standard

- Vendor-neutral telemetry



- Context-based logging/tracing
- Data pipeline from data generation to visualization
- Connects well with visualization platforms such as Kibana and Grafana

An AIOps management system standard

- A reference framework to help organizations implement an AIOps solution by addressing the governance, people, process, and technology aspects.

- It provides a systematic approach to achieving organizational goals
  - ensuring that resources are utilized optimally and that activities are aligned with the organization's objectives.

Dimensions of an AIOpsMS standard

Mission, vision, goals and objectives, capability assessment
strategic alignment, KPIs, etc.

Governance

People	**Process**	**Technology**
Roles & responsibilities (observability specialists, engineers, etc.) Training needs	Organizational processes for the operations of AIOps solution Processes for compliance and controls	AI models Tools & platforms Telemetry standards, Etc.

Continuous Improvement Culture

Guidelines & Best Practices

Maturity Level Assessment

Conclusion

1	Companies must implement an AIOps solution to manage the complexity of today’s IT infrastructures.
2	AIOps research is a growing fields that ranges from data management to root cause analysis and mitigation and anomaly detection
3	Future development of AIOps solutions requires a standardized approach to foster innovation, while managing risks.
4	Observability by design and the proposed dimensions (governance, people, process, and technology) of a standard for AIOps can be a solution

Contact Information:

Prof. Wahab Hamou-Lhadj

Department of Electrical and Computer Engineering
Concordia Applied AI Institute
Concordia University, Montreal, Canada

wahab.hamou-lhadj@concordia.ca