

# From Time Series Foundation Model to Reasoning and Agentic Systems

*A Roadmap for Emergent Anomaly Detection Intelligence*

Chenghao Liu  
Datadog AI Research  
12/7/2025



DATADOG

# Agenda

---

**01** Anomaly Detection & Ops Intelligence

---

**02** TimeRCD: TSFM for Anomaly Detection

---

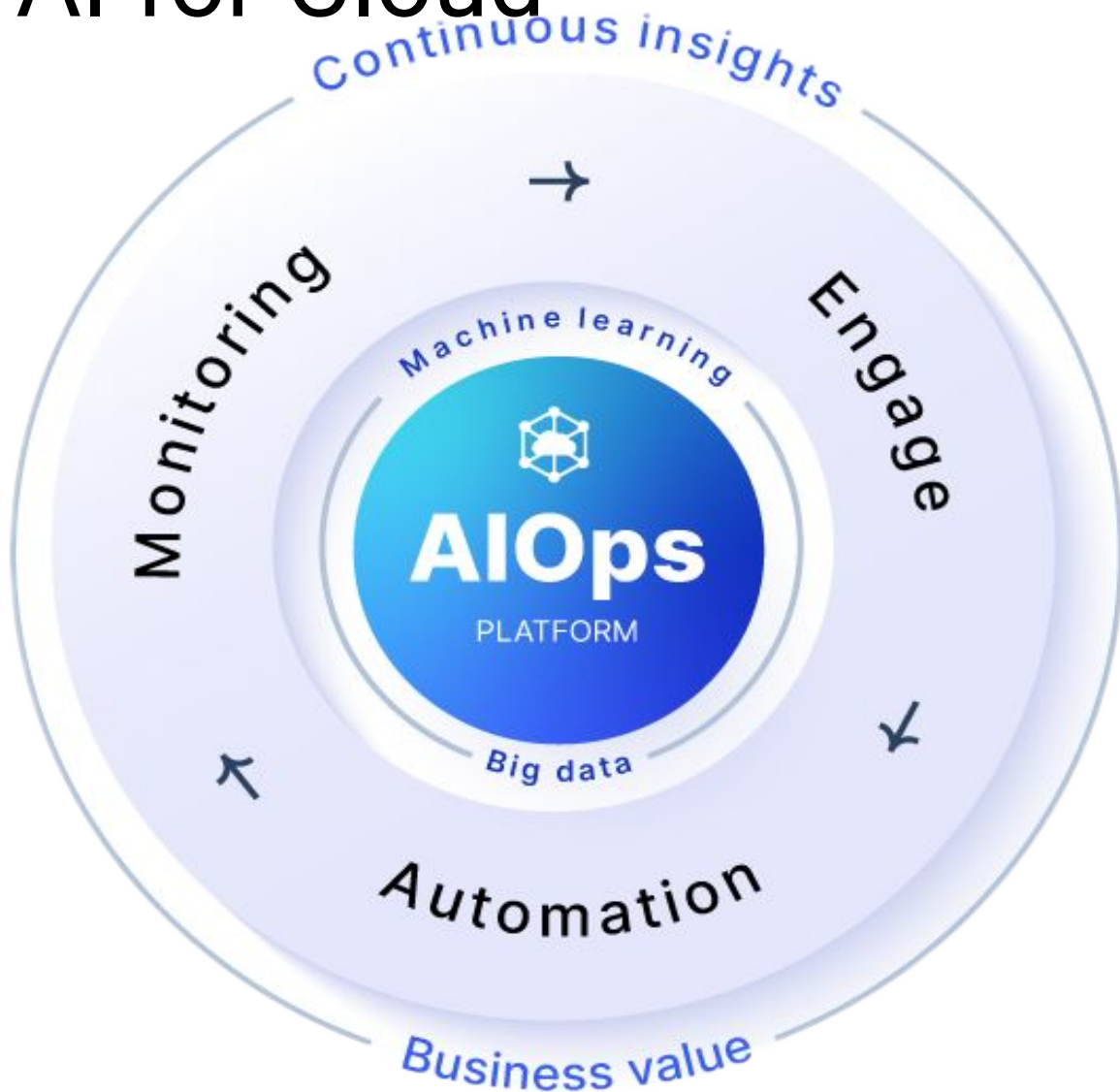
**03** Axis: chat-TSFM for Semantic Explanation

---

**04** Ops Agent & Future Work

# Ops Intelligence

## AI for Cloud



- **Data:** metrics, log, trace
- **Monitoring**  
proactive/reactive incident detection
- **Engage**  
root cause analysis
- **Automation**  
code repairment  
autoscaling  
resource management

# Ops Intelligence

## AI for Healthcare



**Monitor:**  
time series anomaly  
detection  
time series classification



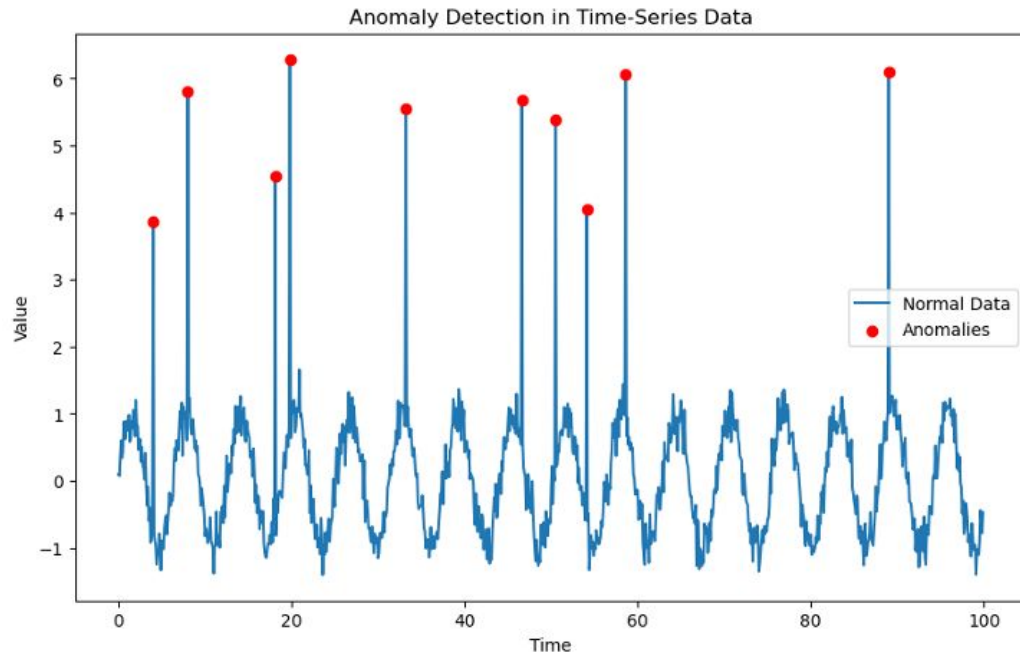
**Diagnosis:**  
root cause analysis  
semantic explainability



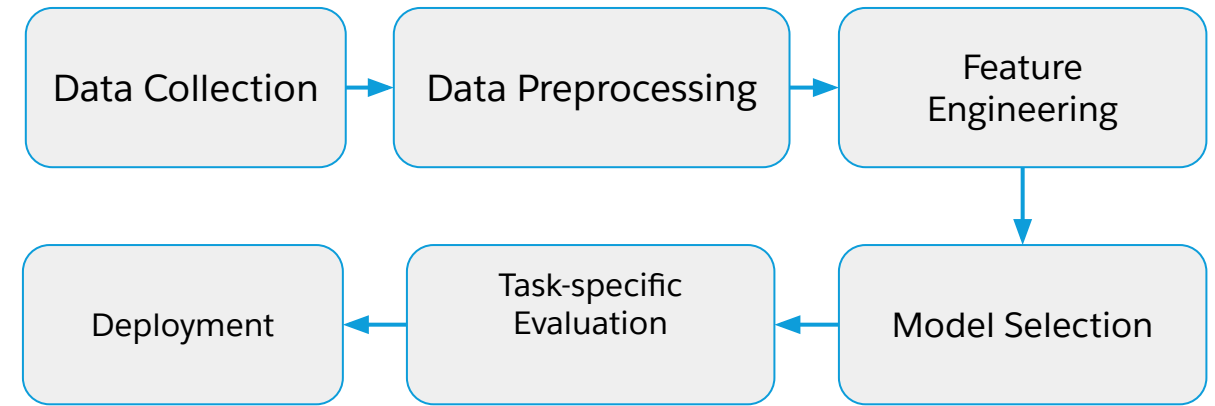
**Automation:**  
dynamic treatment  
recommendation

source: [www.thetatechnolabs.in/blog-posts/](http://www.thetatechnolabs.in/blog-posts/)

# Time Series Anomaly Detection



- Classic Models: iForest, LOF
- Deep Learning Models: OmniAnomaly, TranAD



**One model per application !**

# The Crisis of Traditional TSAD

- **Detection models must be MORE GENERALIZABLE**
  - trained narrowly on domain-specific data, causing them to capture dataset-specific features rather than universal failure patterns .
  - require expensive, exhaustive retraining for new environments or pattern shifts.

Build a Foundation Model, supporting zero-shot TSAD !



# The Crisis of Traditional TSAD

- **Detection systems must be MORE EXPLAINABLE**
  - **Trust and Actionability:** Semantic explainability is vital for precise Diagnosis (RCA) and building Trust with the end-user
  - **Cross-Modal Fusion & Accuracy:** Utilizing text as a semantic fusion center enables the incorporation of other modal information (e.g., logs, traces), high detection accuracy .

Develop a chat-TSFM, providing semantic explanation (linguistic reasoning)!

# The Crisis of Traditional TSAD

- **Ops (Detection) systems must be MORE AUTONOMOUS**
  - Closing the Ops loop: transition from providing alerts to enabling self-correction and automated intervention (resolution) within the ops system's closed-loop workflow.
- *Build an Ops Agent, enabling closed-loop analysis and reliable real-world intervention!*

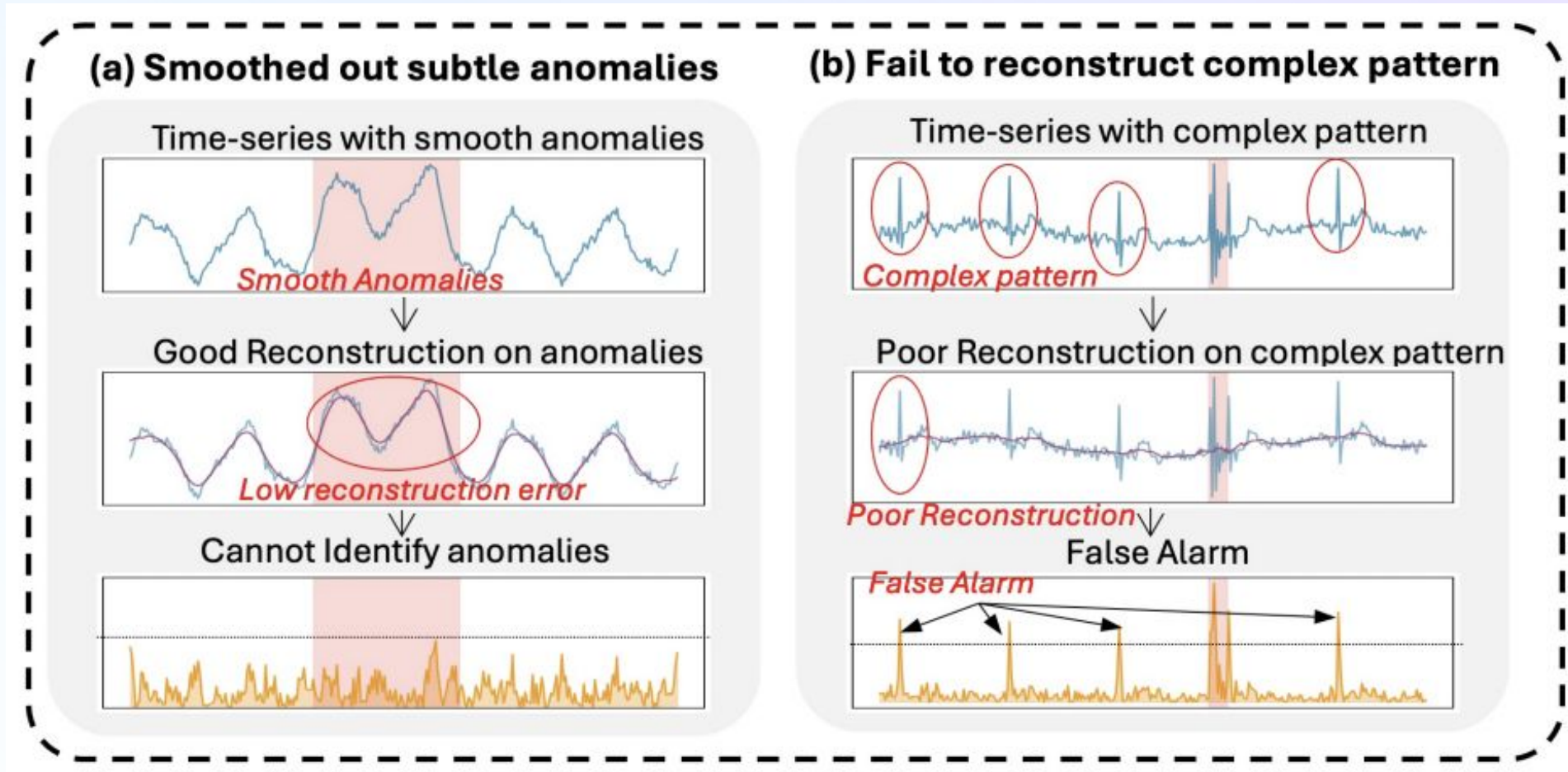


# TimeRCD: A Foundation Models for Zero-Shot Time Series Anomaly Detection

# Challenges & Limitations

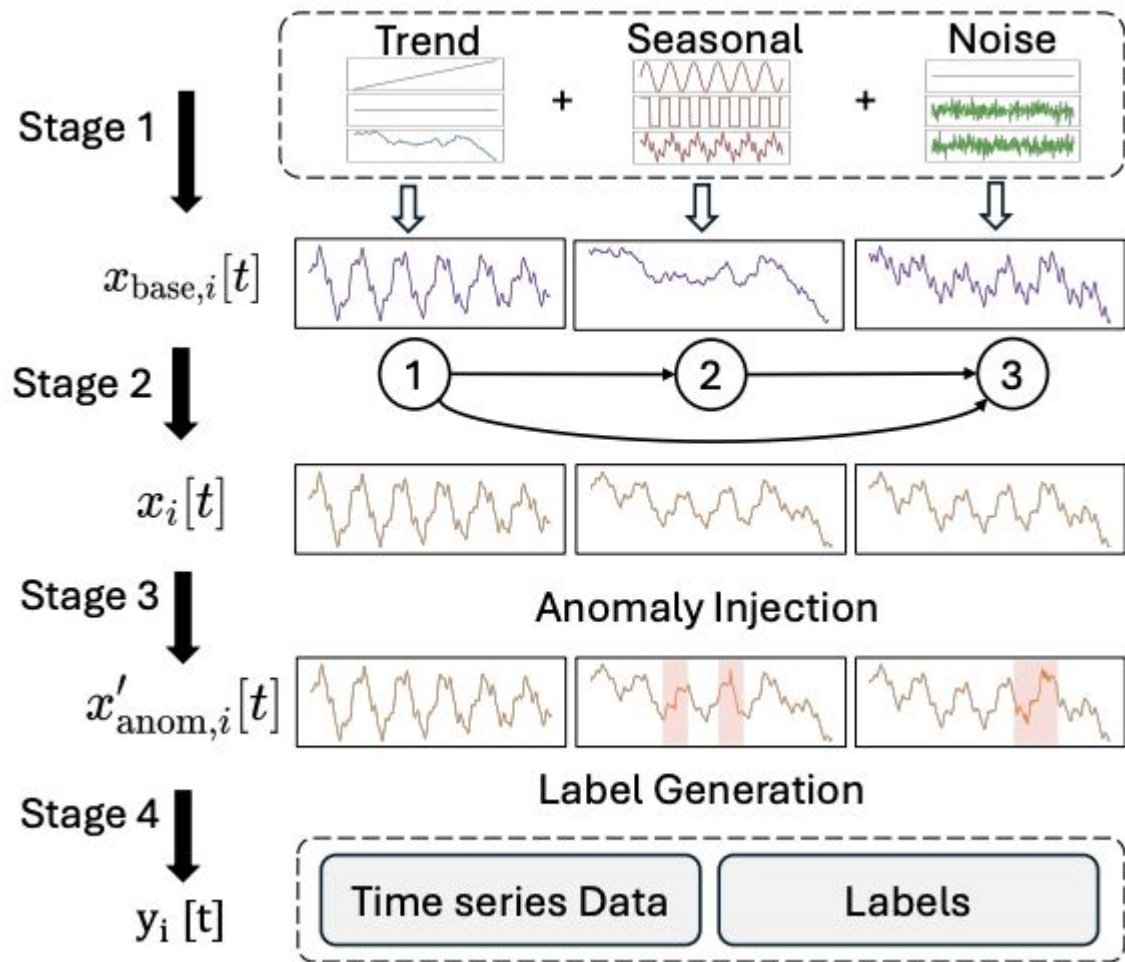
- Pre-training Data:
  - labeled anomalies are inherently **scarce**, providing the model with few examples to learn abnormal behavior.
  - data often lacks **diversity**, covering only a subset of real-world patterns.

# Challenges & Limitations



Reconstruction-based loss with unsupervised learning suffer from **Objective Mismatch**

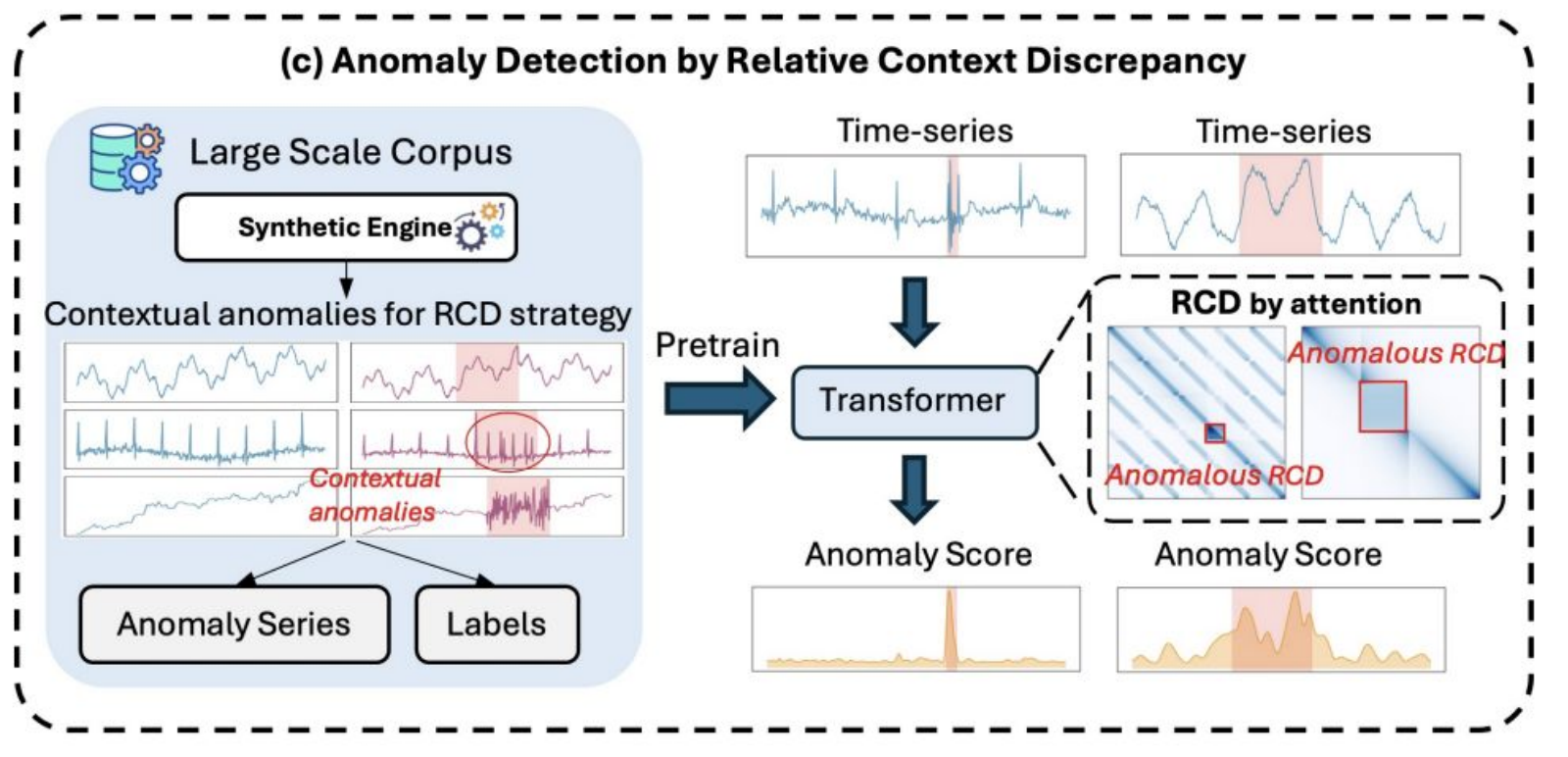
# TimeRCD



## Pretraining Data Recipe:

1. Context-Template Generation
2. Joint-Context Fusion
3. Causal-contextual Anomaly Injection
4. Labels and Masks

# TimeRCD



**Relative Context Discrepancy:** formulate detection as comparing a set of time windows to extract extract discriminative relational patterns.



# Experiments: zero-shot

Metric	Model	Univariate Datasets											Multivariate Datasets					Total 1st	Total 2nd
		IOPS	MGAB	NAB	NEK	Power	SED	Stock	TODS	UCR	WSD	YAHOO	MSL	PSM	SMAP	SMD	SWaT		
<b>Zero-Shot Models</b>																			
Affiliation-F	TimeRCD	<b>83.28</b>	<b>70.69</b>	82.48	79.73	<b>85.51</b>	<b>96.87</b>	71.84	86.37	<b>84.63</b>	90.33	<b>96.65</b>	<b>81.16</b>	<b>81.61</b>	<b>87.73</b>	<b>92.58</b>	<b>71.55</b>	09	02
	DADA <sup>†</sup>	89.37*	67.66*	<b>86.56</b>	<b>95.40</b>	69.79	65.18	<b>98.77</b>	76.89	72.21	<b>93.92</b>	92.20*	<b>76.57</b>	<b>81.27</b>	<b>76.92</b>	<b>83.74</b>	<b>76.18</b>	03	06
	TS-Pulse	68.76	67.33	70.80	73.05	69.94	67.44	67.93	67.90	67.70	68.22	70.05	70.14	70.28	69.21	68.21	71.18	00	00
	MOMENT <sup>†</sup>	87.54*	66.76*	90.45*	92.26	<b>75.97</b>	59.13	45.26	59.76	<b>75.77</b>	<b>95.39</b>	79.99*	74.55*	65.79	77.42*	74.00*	70.17	01	02
	TimesFM	81.88	66.95	79.73	90.49	69.88	67.14	<b>97.53</b>	<b>89.08</b>	70.03	78.97	91.28	20.35	71.24	45.44	62.85	44.37	00	02
	Chronos	<b>90.12</b>	<b>67.89</b>	<b>86.66</b>	<b>93.63</b>	69.72	<b>67.89</b>	96.85	<b>91.96</b>	74.35	90.98	<b>96.34</b>	75.52	70.88	72.22	75.31	70.43	03	04
	Time MOE	76.34	67.23	80.51	80.50	71.19	60.98	63.28	54.68	73.56	80.25	69.70	69.85	54.74	74.38	69.97	64.37	00	00
F1-T	TimeRCD	28.44	<b>1.81</b>	<b>38.85</b>	35.87	<b>28.47</b>	<b>69.43</b>	31.73	<b>65.89</b>	<b>34.30</b>	35.04	<b>85.86</b>	<b>42.47</b>	<b>37.98</b>	<b>33.74</b>	<b>53.91</b>	<b>30.28</b>	11	01
	DADA <sup>†</sup>	42.50*	0.91*	37.24	<b>47.98</b>	19.80	9.56	<b>95.49</b>	35.18	7.22	<b>48.46</b>	79.52*	<b>34.58</b>	<b>31.84</b>	<b>30.42</b>	<b>40.80</b>	<b>35.13</b>	03	05
	TS-Pulse	4.10	0.81	34.61	27.07	19.90	9.71	15.98	13.45	5.12	4.57	5.50	23.57	25.39	12.34	9.15	28.58	00	00
	MOMENT <sup>†</sup>	33.15*	0.80*	52.27*	<b>63.66</b>	<b>19.91</b>	9.54	18.04	17.47	<b>13.02</b>	<b>41.98</b>	11.69*	25.97*	27.77	17.93*	28.68*	28.76	01	03
	TimesFM	<b>48.95</b>	0.93	36.74	36.63	19.80	9.58	88.94	51.13	10.78	41.38	<b>83.46</b>	7.83	25.42	11.64	18.65	21.39	01	01
	Chronos	<b>45.45</b>	<b>1.10</b>	36.10	33.16	19.90	13.18	<b>89.30</b>	<b>53.90</b>	10.88	39.82	<b>79.00</b>	15.59	25.42	11.72	17.32	28.88	00	04
	Time MOE	25.95	0.63	<b>38.70</b>	15.78	19.85	<b>17.73</b>	34.13	20.91	8.29	22.60	37.11	23.92	26.82	14.22	19.90	30.11	00	02
Standard-F1	TimeRCD	24.22	<b>1.62</b>	<b>27.70</b>	33.05	<b>28.59</b>	<b>69.88</b>	32.61	<b>67.02</b>	<b>28.13</b>	31.96	<b>87.02</b>	<b>30.66</b>	<b>26.00</b>	<b>30.48</b>	<b>44.89</b>	<b>28.73</b>	11	01
	DADA <sup>†</sup>	32.76*	0.80*	<b>26.91</b>	<b>48.24</b>	15.99	2.69	<b>95.59</b>	28.18	3.36	<b>45.06</b>	79.30*	<b>22.13</b>	24.07	<b>26.75</b>	<b>34.98</b>	<b>34.78</b>	03	05
	TS-Pulse	3.54	0.73	21.61	23.96	<b>18.27</b>	8.84	15.46	12.45	2.05	2.17	4.00	12.56	22.31	7.44	8.00	23.84	00	01
	MOMENT <sup>†</sup>	30.69*	0.67*	44.75*	<b>63.85</b>	16.39	3.36	19.38	14.64	<b>9.00</b>	<b>41.42</b>	10.54*	14.43*	23.83	12.92*	29.78*	21.30	01	02
	TimesFM	<b>34.28</b>	0.83	26.46	38.15	16.73	2.96	89.13	40.08	7.86	38.50	<b>84.44</b>	5.75	22.18	10.46	18.65	22.84	01	01
	Chronos	<b>32.69</b>	<b>0.99</b>	26.22	33.54	17.47	8.74	<b>89.41</b>	<b>40.52</b>	8.21	34.58	78.89	11.63	22.27	9.62	17.50	24.03	00	04
	Time MOE	26.52	0.45	26.20	11.47	12.16	<b>17.73</b>	34.32	16.38	4.09	20.09	27.50	12.85	<b>24.80</b>	9.01	21.62	23.58	00	02
VUS-PR	TimeRCD	<b>20.23</b>	<b>1.05</b>	<b>24.32</b>	27.88	<b>21.25</b>	<b>80.75</b>	77.28	<b>93.46</b>	<b>23.09</b>	21.77	<b>84.41</b>	<b>20.45</b>	<b>18.69</b>	<b>22.68</b>	<b>37.03</b>	17.58	10	02
	DADA <sup>†</sup>	24.97*	0.57*	<b>24.73</b>	<b>46.85</b>	10.61	6.42	<b>99.51</b>	64.83	2.94	<b>33.42</b>	70.74*	<b>12.74</b>	<b>17.17</b>	<b>20.02</b>	<b>25.98</b>	<b>21.13</b>	03	06
	TS-Pulse	4.64	0.56	16.40	19.39	<b>11.72</b>	9.11	70.95	45.86	1.20	1.83	9.93	7.41	14.48	3.99	4.56	15.67	00	01
	MOMENT <sup>†</sup>	37.35*	0.56*	45.38*	<b>67.74</b>	10.50	4.31	76.97	56.45	6.17	<b>55.26</b>	30.81*	9.32*	16.48	8.97*	15.96*	14.90	02	00
	TimesFM	<b>19.56</b>	0.58	24.01	35.02	10.44	6.13	<b>98.39</b>	<b>72.89</b>	6.03	21.57	<b>86.78</b>	11.84	14.76	16.95	13.02	<b>19.43</b>	01	04
	Chronos	19.00	<b>0.60</b>	23.76	31.80	10.95	8.65	97.49	70.66	<b>6.56</b>	18.81	83.54	8.25	14.61	5.18	10.22	16.44	00	02
	Time MOE	16.63	0.52	22.62	19.76	9.34	<b>10.87</b>	74.78	48.78	2.10	10.93	20.90	7.82	15.68	4.98	11.12	16.20	00	01
<b>TimeRCD Grand Total (Zero-Shot)</b>																	41	06	

zs-TimeRCD ranks first in 41/64 and second in 6/64 in zero-shot comparisons!



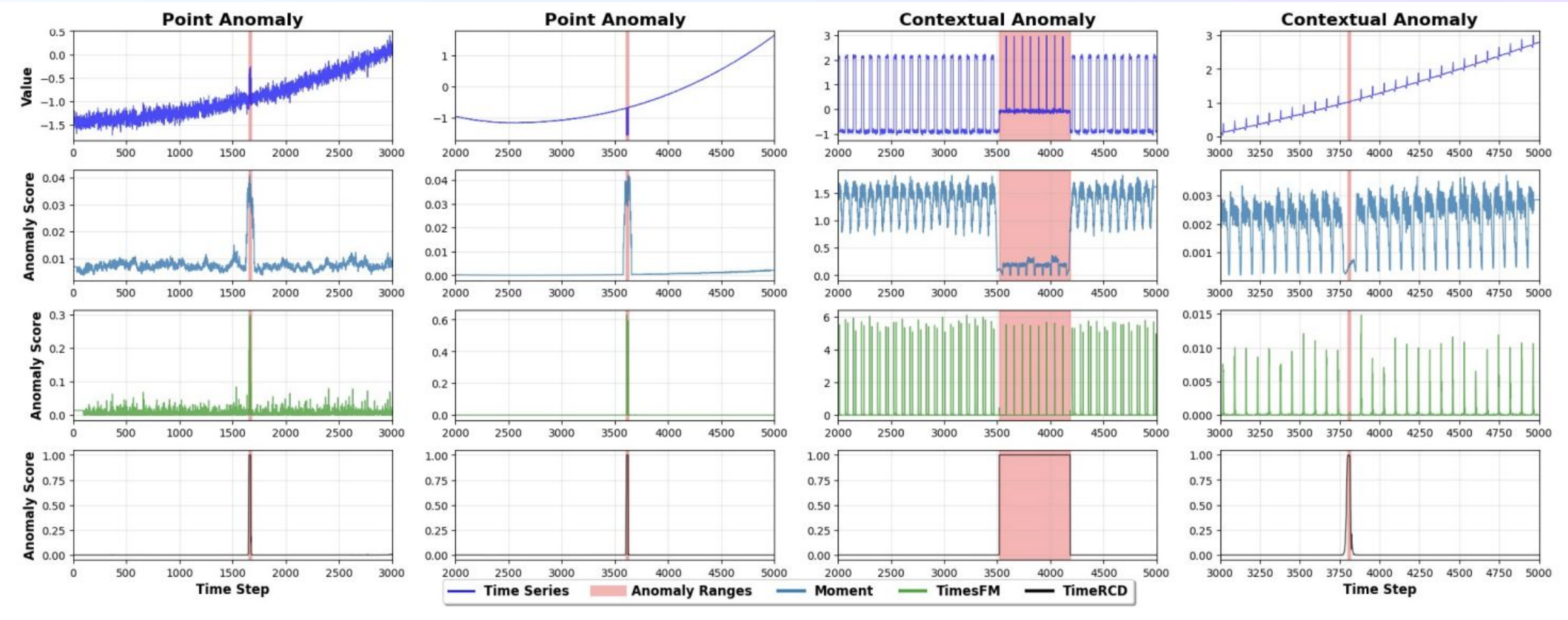
# Experiments: full-shot

		Full-Shot Models																	
Affiliation-F	TimeRCD	<b>83.28</b>	<b>70.69</b>	82.48	79.73	<b>85.51</b>	<b>96.87</b>	<b>71.84</b>	<b>86.37</b>	<b>84.63</b>	<b>90.33</b>	<b>96.65</b>	81.16	<b>81.61</b>	87.73	<b>92.58</b>	71.55	10	01
	TranAD	<b>83.19</b>	67.28	90.28	85.02	71.56	61.03	57.94	52.76	73.31	<b>84.34</b>	76.08	79.91	<b>73.83</b>	87.39	<b>92.20</b>	75.37	00	04
	USAD	71.08	67.81	<b>91.54</b>	71.13	76.48	55.60	35.92	47.90	76.00	65.10	53.05	81.86	<b>57.86</b>	87.25	85.09	75.06	00	01
	OmniAnomaly	80.32	67.35	<b>92.35</b>	<b>86.30</b>	<b>78.16</b>	61.26	<b>75.24</b>	50.73	73.53	78.02	71.31	83.15	58.17	<b>91.38</b>	85.82	73.39	03	02
	LOF	81.06	68.44	75.75	84.74	66.76	63.85	69.74	60.58	73.53	81.29	75.63	<b>84.35</b>	61.98	63.32	64.13	56.34	01	00
	IForest	52.81	<b>68.82</b>	39.84	71.15	0.00	70.09	0.06	44.17	50.56	41.24	33.30	63.36	63.78	59.96	69.71	0.00	00	01
	Sub-PCA	75.39	66.90	89.29	<b>97.10</b>	71.37	67.14	70.63	<b>72.75</b>	<b>76.66</b>	76.45	75.85	<b>84.25</b>	71.49	<b>90.08</b>	85.80	<b>76.29</b>	02	04
	DCdetector	71.83	67.91	72.21	62.31	69.75	<b>72.20</b>	55.79	57.81	70.18	72.79	67.77	67.74	67.32	67.10	69.55	71.07	00	01
TFMAE	78.25	67.50	75.99	76.91	70.30	68.17	56.39	62.83	70.60	80.25	<b>76.87</b>	75.70	70.07	75.36	70.85	<b>75.72</b>	00	02	
F1-T	TimeRCD	28.44	<b>1.81</b>	38.85	35.87	<b>28.47</b>	<b>69.43</b>	31.73	<b>65.89</b>	<b>34.30</b>	<b>35.04</b>	<b>85.86</b>	42.47	<b>37.98</b>	33.74	<b>53.91</b>	30.28	08	01
	TranAD	22.63	1.65	37.28	69.97	22.36	9.57	16.73	13.51	7.75	20.94	8.41	39.42	25.49	29.12	37.98	<b>49.58</b>	00	01
	USAD	20.99	<b>4.07</b>	<b>61.46</b>	70.64	<b>28.23</b>	9.54	16.86	20.85	14.63	14.18	9.35	48.71	28.96	43.94	50.41	<b>50.41</b>	03	01
	OmniAnomaly	<b>51.17</b>	1.61	40.09	<b>82.20</b>	23.48	9.68	<b>36.22</b>	14.33	8.47	<b>34.79</b>	24.16	<b>49.36</b>	<b>30.42</b>	<b>46.63</b>	51.84	46.64	03	04
	LOF	27.97	1.15	35.76	63.57	19.80	9.60	<b>66.14</b>	<b>31.63</b>	8.31	24.38	<b>55.93</b>	38.97	25.58	21.81	10.13	30.62	01	02
	IForest	7.64	0.84	21.44	65.56	0.00	9.54	1.10	11.06	6.36	4.28	4.90	20.73	25.39	14.32	16.20	0.00	00	00
	Sub-PCA	<b>32.75</b>	0.98	<b>54.15</b>	<b>84.18</b>	20.30	9.54	20.10	18.40	<b>18.90</b>	24.62	11.57	<b>49.02</b>	30.38	<b>44.31</b>	<b>51.87</b>	46.65	01	06
	DCdetector	6.61	1.32	32.72	29.21	21.13	10.53	16.07	16.40	6.62	7.32	6.81	23.24	25.34	15.73	9.47	28.64	00	00
TFMAE	19.41	1.07	33.04	31.11	20.18	<b>11.82</b>	22.15	16.48	5.90	19.11	23.85	25.28	25.36	19.39	10.13	28.46	00	01	
Standard-F1	TimeRCD	24.22	<b>1.62</b>	27.70	33.05	<b>28.59</b>	<b>69.88</b>	32.61	<b>67.02</b>	<b>28.13</b>	<b>31.96</b>	<b>87.02</b>	30.66	26.00	30.48	44.89	28.73	06	01
	TranAD	<b>34.85</b>	1.46	27.33	60.36	22.36	2.63	16.23	11.94	4.40	20.23	5.70	29.60	25.63	25.11	43.99	<b>61.86</b>	00	02
	USAD	30.66	<b>3.89</b>	<b>56.15</b>	62.91	<b>28.24</b>	3.41	17.99	23.87	10.74	13.20	7.21	<b>38.71</b>	28.41	38.66	53.06	<b>62.82</b>	03	02
	OmniAnomaly	<b>47.05</b>	1.44	28.81	<b>74.03</b>	23.50	0.43	<b>38.59</b>	12.65	5.11	<b>29.57</b>	21.40	<b>39.10</b>	<b>30.43</b>	<b>40.50</b>	<b>57.06</b>	55.93	04	04
	LOF	30.28	1.05	24.04	56.92	12.18	4.11	<b>66.20</b>	<b>25.77</b>	4.70	22.62	<b>48.95</b>	30.65	18.80	18.70	8.41	29.08	01	02
	IForest	8.37	0.73	29.41	58.10	19.77	3.81	16.91	13.35	4.09	2.07	3.20	14.68	24.15	13.61	16.89	26.76	00	00
	Sub-PCA	33.96	0.83	<b>46.71</b>	<b>85.43</b>	16.05	9.56	21.77	18.72	<b>15.12</b>	24.74	11.06	38.29	<b>30.26</b>	<b>38.90</b>	<b>57.22</b>	55.98	02	04
	DCdetector	5.19	1.21	24.02	17.37	21.10	10.54	16.97	17.85	3.18	4.64	4.14	14.08	25.33	10.67	8.99	27.02	00	00
TFMAE	9.48	0.97	23.78	19.74	20.14	<b>11.87</b>	23.32	14.88	2.83	15.53	20.50	15.68	25.39	12.58	9.16	27.08	00	01	
VUS-PR	TimeRCD	20.23	<b>1.05</b>	24.32	27.88	<b>21.25</b>	<b>80.75</b>	77.28	<b>93.46</b>	<b>23.09</b>	<b>21.77</b>	<b>84.41</b>	20.45	<b>18.69</b>	22.68	37.03	17.58	08	00
	TranAD	21.61	0.64	24.82	61.63	13.04	5.75	78.08	47.33	2.25	12.20	25.78	14.78	16.49	13.37	28.34	<b>47.37</b>	01	00
	USAD	16.58	<b>0.75</b>	<b>55.03</b>	58.53	<b>18.68</b>	4.37	74.53	<b>56.36</b>	8.85	10.00	14.15	29.95	17.59	26.37	34.53	<b>44.73</b>	01	04
	OmniAnomaly	<b>25.35</b>	0.64	27.17	<b>74.51</b>	14.32	6.20	<b>91.29</b>	45.55	2.40	16.37	29.26	<b>31.57</b>	<b>18.58</b>	<b>28.07</b>	<b>37.44</b>	42.97	04	03
	LOF	19.43	0.57	21.18	58.52	9.31	6.81	<b>83.07</b>	49.14	2.39	12.85	<b>41.37</b>	24.67	13.58	10.59	4.40	14.50	00	02
	IForest	8.59	0.62	23.57	56.50	11.56	7.71	70.99	46.62	2.88	2.06	10.47	11.29	15.85	7.55	8.88	15.49	00	00
	Sub-PCA	<b>23.02</b>	0.60	<b>46.08</b>	<b>88.91</b>	10.49	3.72	80.86	54.16	<b>12.92</b>	<b>16.41</b>	21.57	<b>31.43</b>	18.52	<b>26.42</b>	<b>37.50</b>	43.02	02	06
	DCdetector	5.83	0.59	16.60	14.03	12.32	9.37	74.16	46.66	1.53	3.23	10.17	7.01	14.49	4.21	4.66	15.04	00	00
TFMAE	5.32	0.64	15.68	17.81	11.90	<b>9.55</b>	73.54	48.79	2.57	5.36	25.93	8.25	14.22	5.76	4.77	15.38	00	01	
<b>TimeRCD Grand Total (Full-Shot)</b>																		32	03

zs-TimeRCD ranks first in 32/64 and second in 2/64 in full-shot comparisons!



# Contextual Understanding



TimeRCD can learn the complex relationships between distant point

# Real Data or Synthetic Data

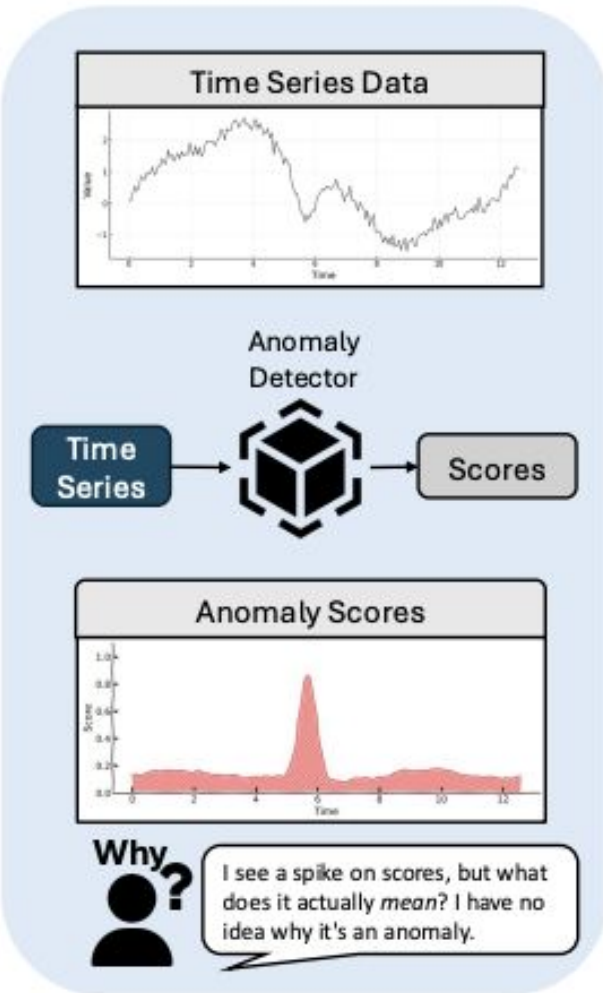
Table 2: Weighted average performance across 9 univariate benchmarks.

Pre-Training Dataset	Affiliation-F	F1-T	Standard-F1	VUS-PR
Our Synthetic Data	<b>0.878</b>	<b>0.569</b>	<b>0.523</b>	0.478
Our Synthetic + DADA Injection	<b>0.878</b>	0.505	0.462	<b>0.487</b>
Real-world Data + DADA Injection	0.716	0.073	0.062	0.102

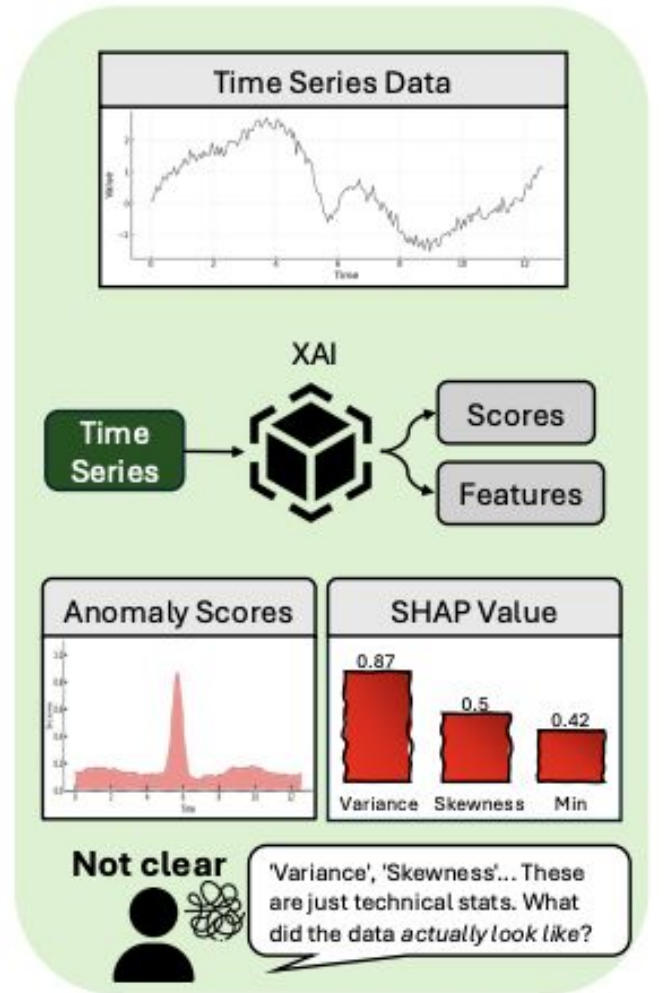
# **AXIS: Explainable Time Series Anomaly Detection with Large Language Models**

# Motivation of Anomaly Explanation

(a) Traditional Black-box Models



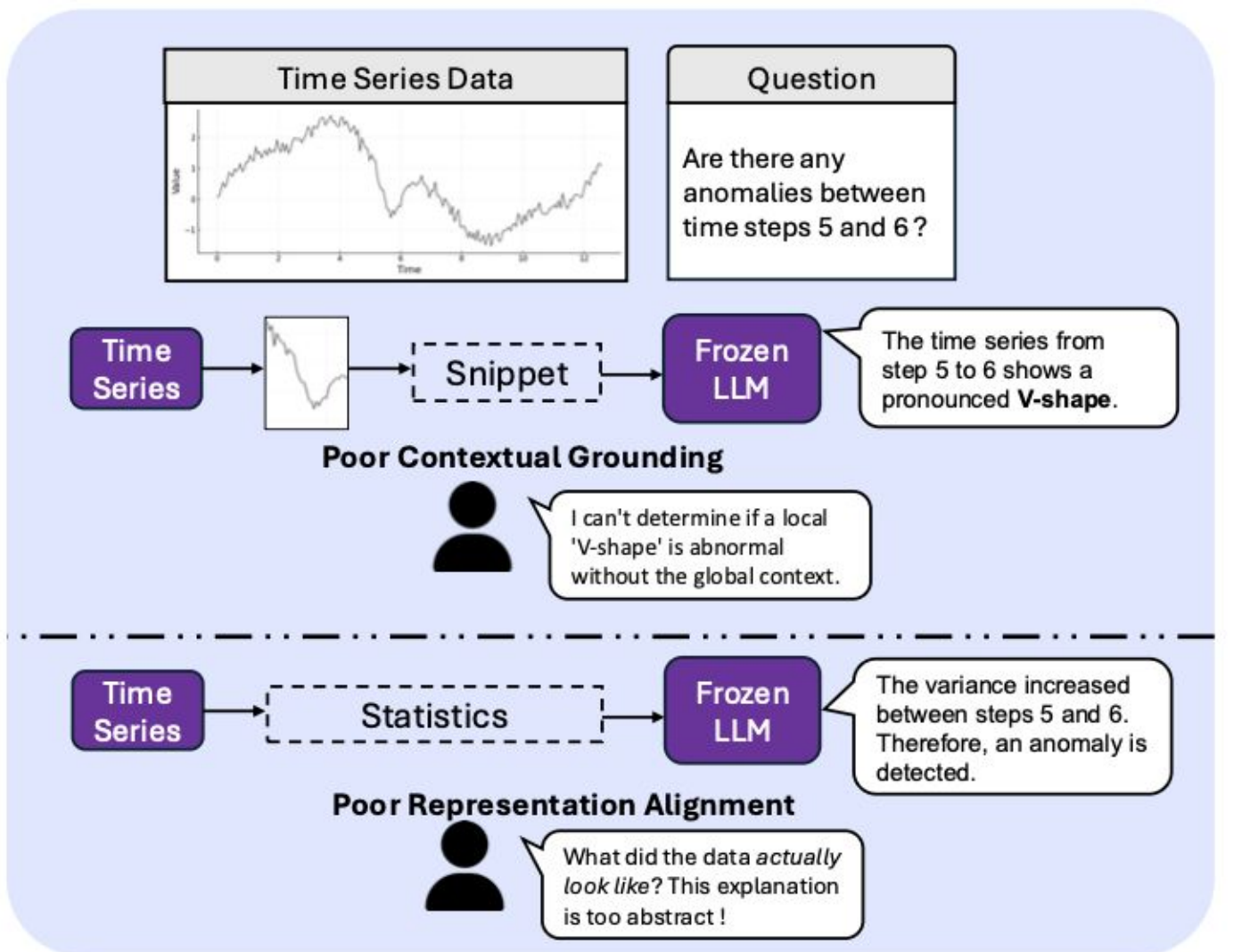
(b) Existing XAI



Reasoning process is a black box. Experts remain in the dark.

Post-hoc attribution methods lack semantic intuitive.

# Limitations



## Lack of Contextual Grounding:

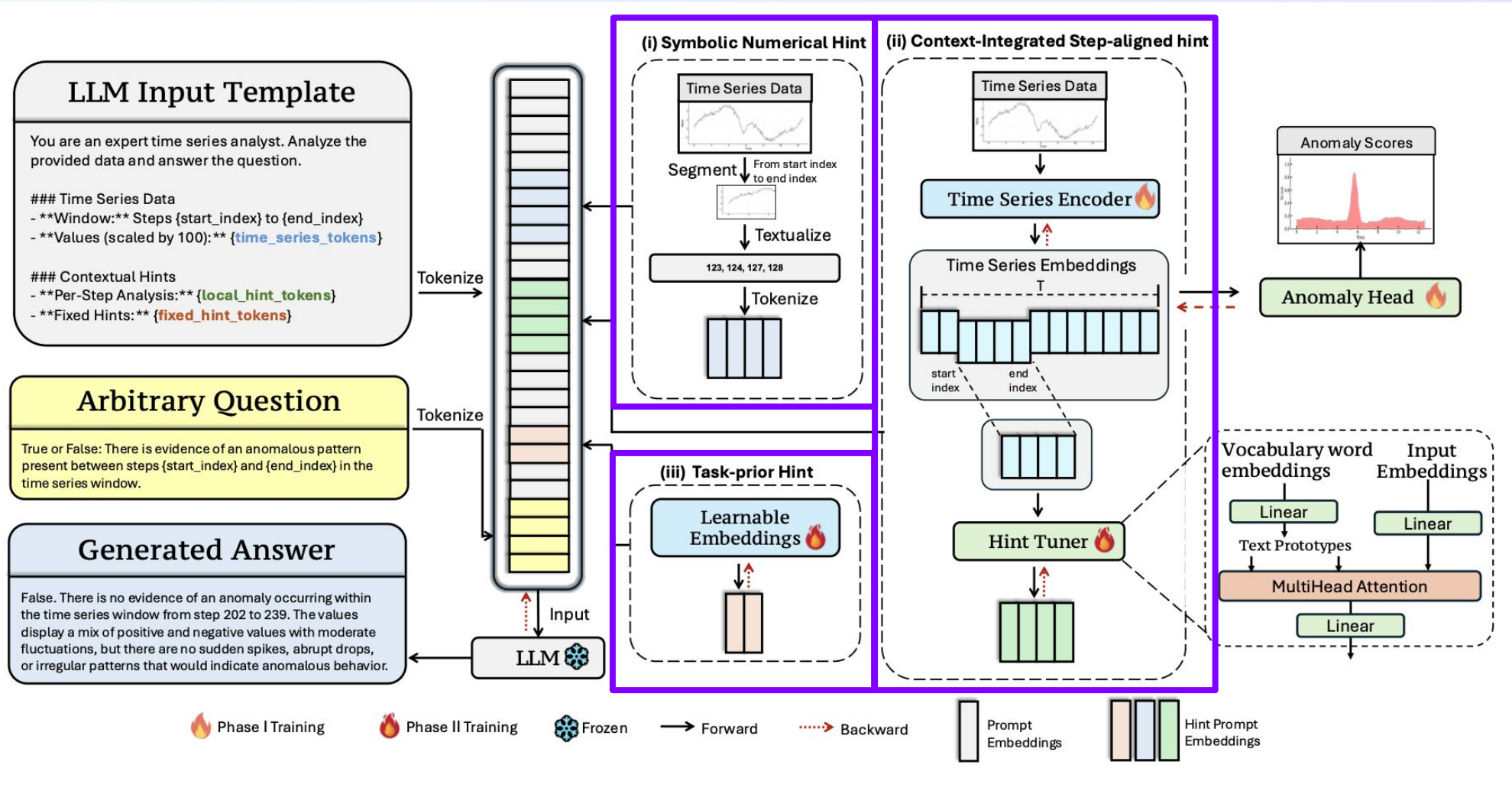
need to interpret local events in the context of the full series

## Lack of Representation Alignment:

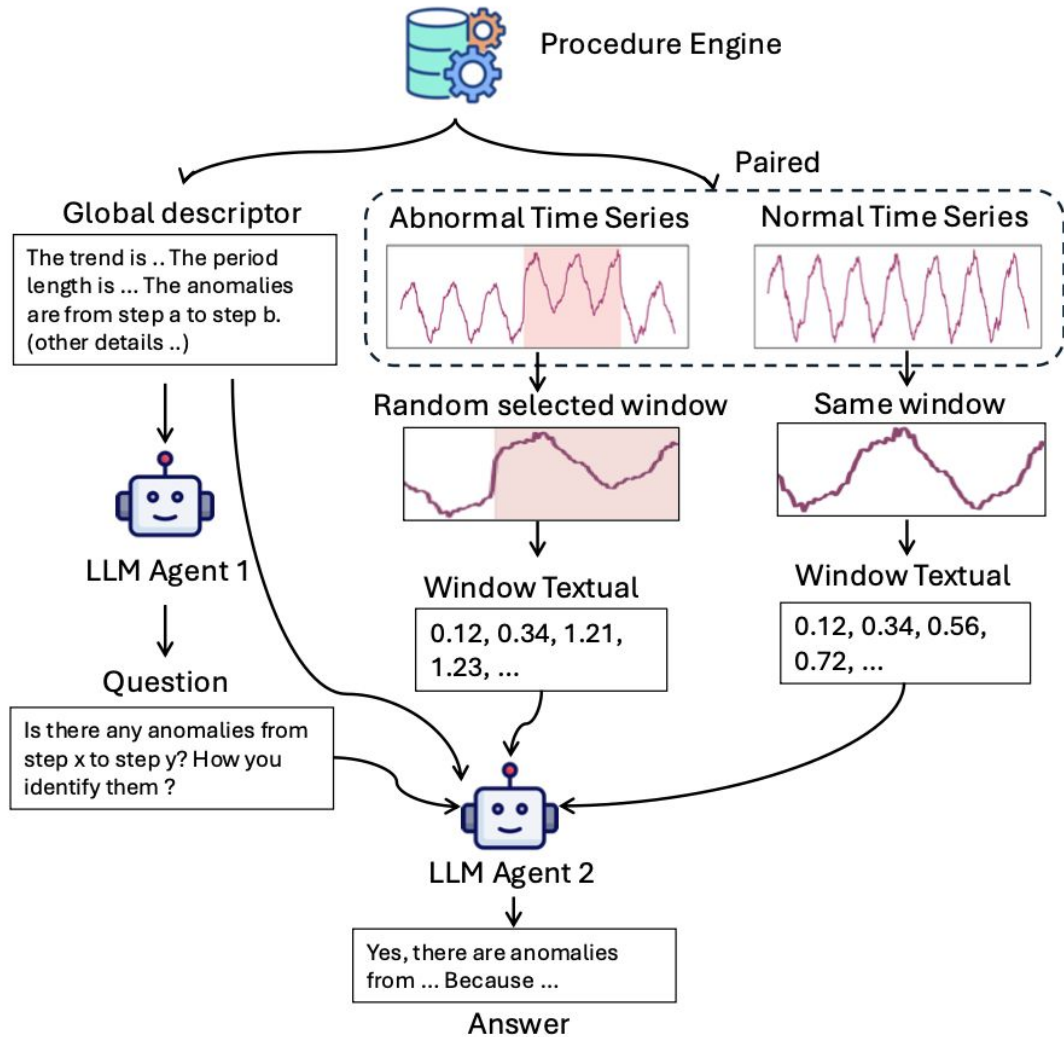
bridge the semantic gap between low-level numerical signals and the rich concepts from human reasoning



# Axis



# Semantic Anomaly Benchmark



1. Pattern-Level Anomaly Vocabulary
2. Contextual and Comparative Reasoning
3. LLM-Powered Explanation Generation
4. Ensuring Benchmark Integrity

Multiple Choice, True/False, Open-Ended QA format.



# Experiments

Evaluation Metric: LLM-as-a-judge approach

Table 1: Main Results: AXIS vs Baselines

Model	Multiple Choice			Open Ended				True False		
	Final	Corr.	Rsn. Qual.	Final	Acc.	Comp.	Rel.	Final	Corr.	Justif.
<b>AXIS</b>	<b>4.19</b>	<b>4.21</b>	<b>4.14</b>	<b>3.02</b>	<b>2.87</b>	<b>2.93</b>	<b>3.31</b>	<b>3.65</b>	<b>3.60</b>	<b>3.74</b>
Image LLM	<u>4.09</u>	<u>4.12</u>	<u>4.02</u>	2.68	2.53	2.49	3.07	2.64	2.57	2.74
ChatTS	3.29	3.40	3.05	2.19	1.67	2.13	2.87	2.79	2.76	2.83
LLMAD	2.73	2.70	2.79	2.09	2.09	1.89	2.31	2.49	2.52	2.43
ChatTime	1.33	1.49	0.98	0.96	0.95	0.98	0.95	1.04	1.07	1.00
AnomLLM(Full)	3.13	2.98	3.49	<u>2.86</u>	2.53	<u>2.89</u>	3.20	2.88	2.60	<u>3.31</u>
AnomLLM(Window)	3.78	3.81	3.70	<u>2.84</u>	<u>2.78</u>	<u>2.55</u>	<u>3.24</u>	<u>3.32</u>	<u>3.45</u>	3.12
Baseline 1 (Random)	1.02	1.03	1.00	1.21	1.21	1.05	1.41	1.29	1.36	1.18
Baseline 2 (Heuristic)	2.44	2.81	1.58	1.72	1.98	1.15	2.07	2.64	2.74	2.50

# Experiments

Performance on real data

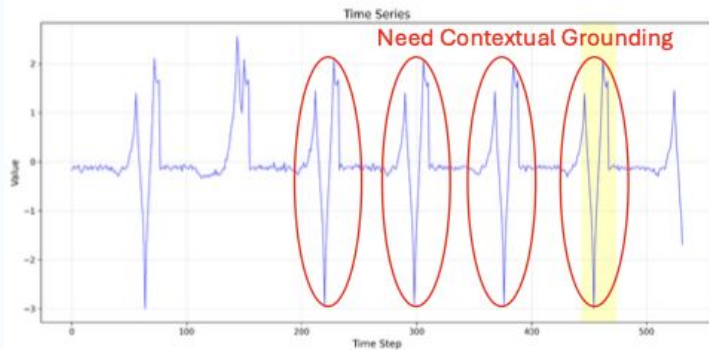
Table 4: Performance comparison on Multiple Choice (MC) and True/False (TF) tasks.

Dataset	Metric	AnomLLM		LLMAD	ChatTS	AXIS 14B	Image LLM
		Window	Full				
YAHOO (W=0.651)	MC Score	2.78	1.69	2.48	2.83	<b>3.07</b>	2.80
	MC Acc.	0.48	0.18	0.47	0.52	<b>0.55</b>	0.52
	TF Score	2.62	1.90	2.54	2.13	<b>3.28</b>	2.29
	TF Acc.	0.53	0.38	0.51	0.45	<b>0.67</b>	0.52
TODS (W=0.672)	MC Score	2.07	1.77	2.33	2.67	<b>2.90</b>	2.53
	MC Acc.	0.37	0.30	0.33	0.47	<b>0.53</b>	0.47
	TF Score	2.67	2.53	2.87	2.53	<b>3.07</b>	2.27
	TF Acc.	0.53	0.52	0.58	0.53	<b>0.62</b>	0.50
NEK (W=0.713)	MC Score	3.31	1.94	1.44	3.06	3.25	<b>3.41</b>
	MC Acc.	<b>0.63</b>	0.31	0.13	0.56	<b>0.63</b>	<b>0.63</b>
	TF Score	2.88	1.69	2.94	2.25	<b>3.69</b>	2.25
	TF Acc.	0.56	0.35	0.59	0.47	<b>0.74</b>	0.50



# Experiments

## (a) Visualization for contextual grounding ability



**Question:** When examining the time series window from step 444 to 473, what evidence would you look for to support or refute the presence of subtle anomalies, and how would you differentiate between normal fluctuations and genuine outlier behavior near the window boundaries?

**Ours (Score 5.00) :** There is **no evidence of anomalies** in the time series window from step 444 to 473. The pattern shows a smooth transition from positive to negative values, followed by a gradual recovery and stabilization. Near the window boundaries, the values remain within a predictable range, indicating normal behavior rather than outlier activity. **Thus, the observed fluctuations are characteristic of typical time series dynamics** and do not suggest the presence of subtle or genuine anomalies in this segment.

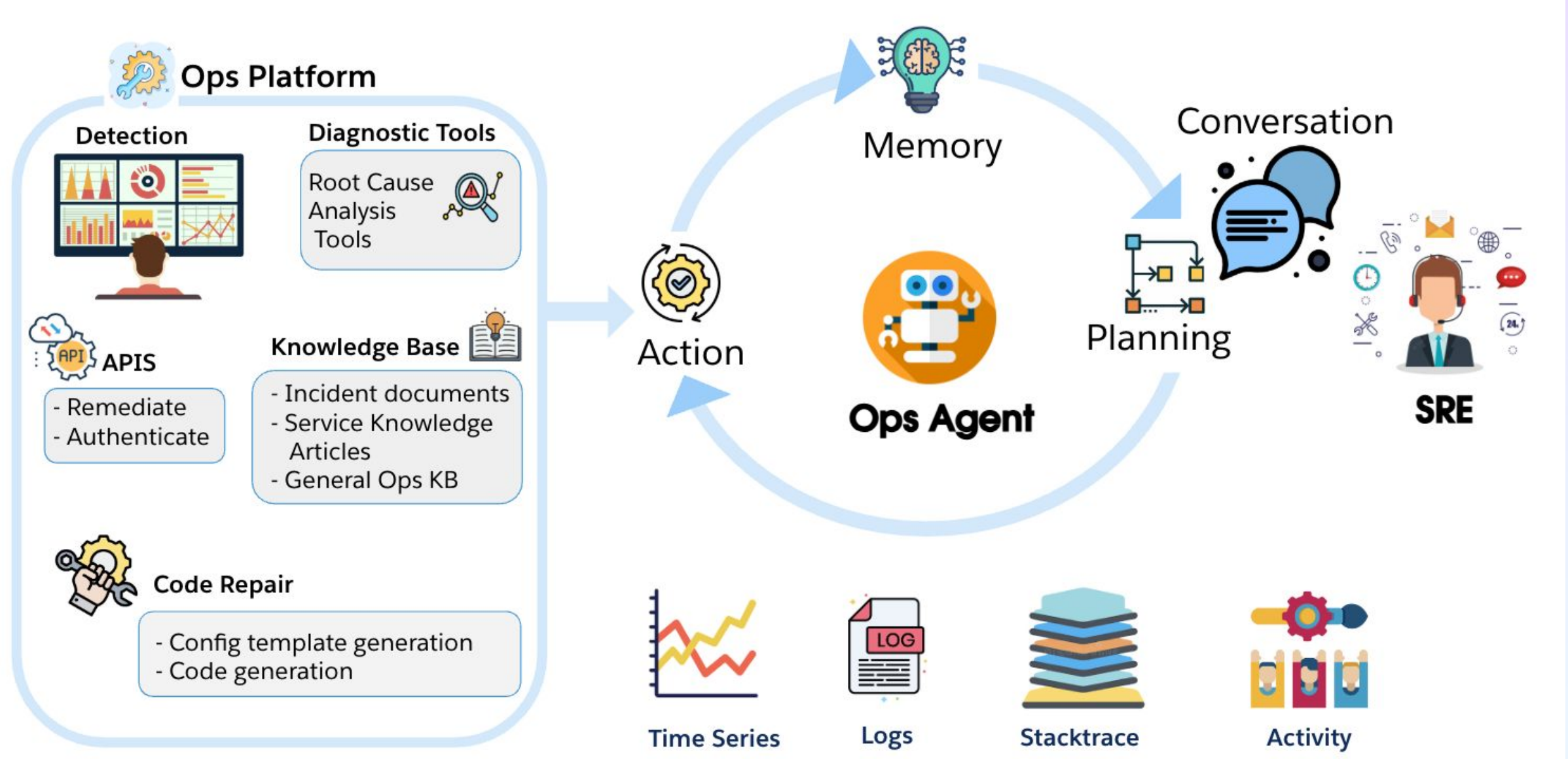
**ChatTS (Score 1.65):** In this window, the primary concern is the transition from the steady trend to the final downward trend. **The periodic fluctuation amplitude is approximately 3.3, with a low noise level (standard deviation of 0.06).** To distinguish between normal fluctuations and outliers, monitor deviations exceeding +2.0 from the baseline. The sudden decrease near the end of the time series (around point 528) may appear as an anomaly in this window.

**AnomLLM (Score 1.65):** To identify subtle anomalies from steps 444 to 473, I would examine deviations beyond typical fluctuation ranges, such as sudden spikes or dips inconsistent with prior patterns. Differentiating normal fluctuations from outliers involves assessing whether boundary points significantly deviate from the local mean or trend, considering the context of surrounding data. Consistent, **isolated deviations near boundaries may suggest genuine outliers**, whereas gradual, minor variations are likely normal fluctuations.

**LLMAD (Score 3.60):** There are no segments with persistent high or low values that deviate significantly from the rest of the sequence. **The brief dip to -2 at index 11 is immediately surrounded by values close to zero and quickly returns to baseline**, which does not constitute a prolonged or abrupt anomaly. Upon careful review, all observed fluctuations are minor and fall within the expected range of normal data. There are no sequences or points that meet the criteria for any anomaly type with high confidence.

# Ops Agent

# Ops Agent





# Datadog AI Research: Observability Intelligence



Looking for Talent!

