TABARD: A Novel Benchmark for Tabular Anomaly Analysis, Reasoning and Detection

Manan Roy Choudhury[1] , Anirudh Iyengar Kaniyar Narayana Iyengar[1] , Shikhhar Siingh[1]* , Sugeeth Puranam [2]*, Vivek Gupta[1]

[1]Arizona State University, [2]University of Utah, *Equal Contribution

WHAT IS IT ABOUT?

What Are ANOMALIES IN TABLES

I. **Context-Dependent Errors:** Require deep contextual understanding and reasoning to be found.

II. **Silent Data Corruptors:** They poison data, leading to flawed decisions and financial loss.

III. **More Than Outliers:** Often subtle and complex, hiding beyond simple statistical checks.

IV. **Rule Breakers:** They defy traditional rule-based and statistical detection methods.

V. **Trust Eroding:** Undetected anomalies compromise trust in data across all industries.

"RARE but not always WRONG"

WHY THIS MATTERS?

Anomaly Detection in Tables

The Problems:

- **Decisions at Risk**: Undetected table anomalies corrupt data, leading to flawed decisions, financial losses, and compromised trust across industries.
- **Subtle & Complex Errors**: Anomalies are often diverse, subtle, and require deep contextual understanding beyond simple statistical outliers.
- **Traditional Tools Fail**: Rule-based and statistical methods are brittle, lack reasoning, and cannot adapt to the complex, semantic nature of many table anomalies.
- **No Human-like Reasoning**: Current approaches cannot interpret context or apply common sense, which is crucial for identifying sophisticated errors.
- **Benchmark Gap**: A lack of comprehensive benchmarks prevents effective evaluation of advanced, reasoning-based anomaly detection techniques.

MOTIVATION

Case Studies

The financial giant Lehman Brothers filed for bankruptcy on Sept. 15, 2008, with $613 billion in debt, putting thousands of employees out of work and sending the already recessionary economy into a tailspin.

Fannie Mae $1.2bn Restatement

Published by Myles Arnott

Mar 17, 2020 12:52:00 PM

Covid: how Excel may have caused loss of 16,000 test results in England

Alex Hern

UK technology editor

Public Health England data error blamed on limitations of Microsoft spreadsheet

Enron Scandal and Accounting Fraud: What Happened?

Case Study 18: How Excel Errors and Risk Oversights Cost JP Morgan $6 Billion

In the spring of 2012, JP Morgan Chase & Co. faced one of the most significant financial debacles in recent history, known as the "London Whale" incident. The debacle resulted in losses amounting to approximately $6 billion, fundamentally shaking the confidence in the bank's risk management practices.

At the core of this catastrophe was the failure of the Synthetic Credit Portfolio Value at Risk (VaR) Model, a sophisticated financial tool intended to manage the risk associated with the bank's trading strategies.

FCA fines HSBC Bank plc £63.9 million for deficient transaction monitoring controls

Press Releases | First published: 17/12/2021 | Last updated: 06/05/2022 | See all updates

Print Page

Share page

The FCA has fined HSBC Bank plc (HSBC) £63,946,800 for failings in its anti-money laundering processes.

CORAL

Anomalous Table

Data Consistency Anomaly
(Duplicate Primary Key)

Temopral Anomaly
(Incorrect Time context)

Factual Anomaly
(Imaginary City)

Order ID	Dates (Order / Ship)	Transaction_Details (Item, Qty, Disc (%), Price ($), Total($)))	Card Info	Locale
1001	2025-05-12/2025-05-20	("Laptop", 2, "N/A", 1200, 2000)	4111 xxxx xxxx xxxx	USA
1001	1600-01-01/1600-01-15	("Time Machine", 5, 150, 1000, 5000)	5500 0000 0000 0004	Atlantis
1002	2025-05-10/2025-05-20	("Electric Scooter", -50, "N/A", 500, 0)	6011 xxxx xxxx xxxx	Canada
1003	2025-04-20/2025-04-18	("Snow Boots", 1, "N/A", 80, 80)	3782 8224 6310 0050	Singapore

Logical Anomaly
(Ship Date before Order Date)

Value Anomaly
(Negative Value)

Calculation Anomaly
(Incorrect Total)

Security Anomaly
(Non Encrypted Card Details)

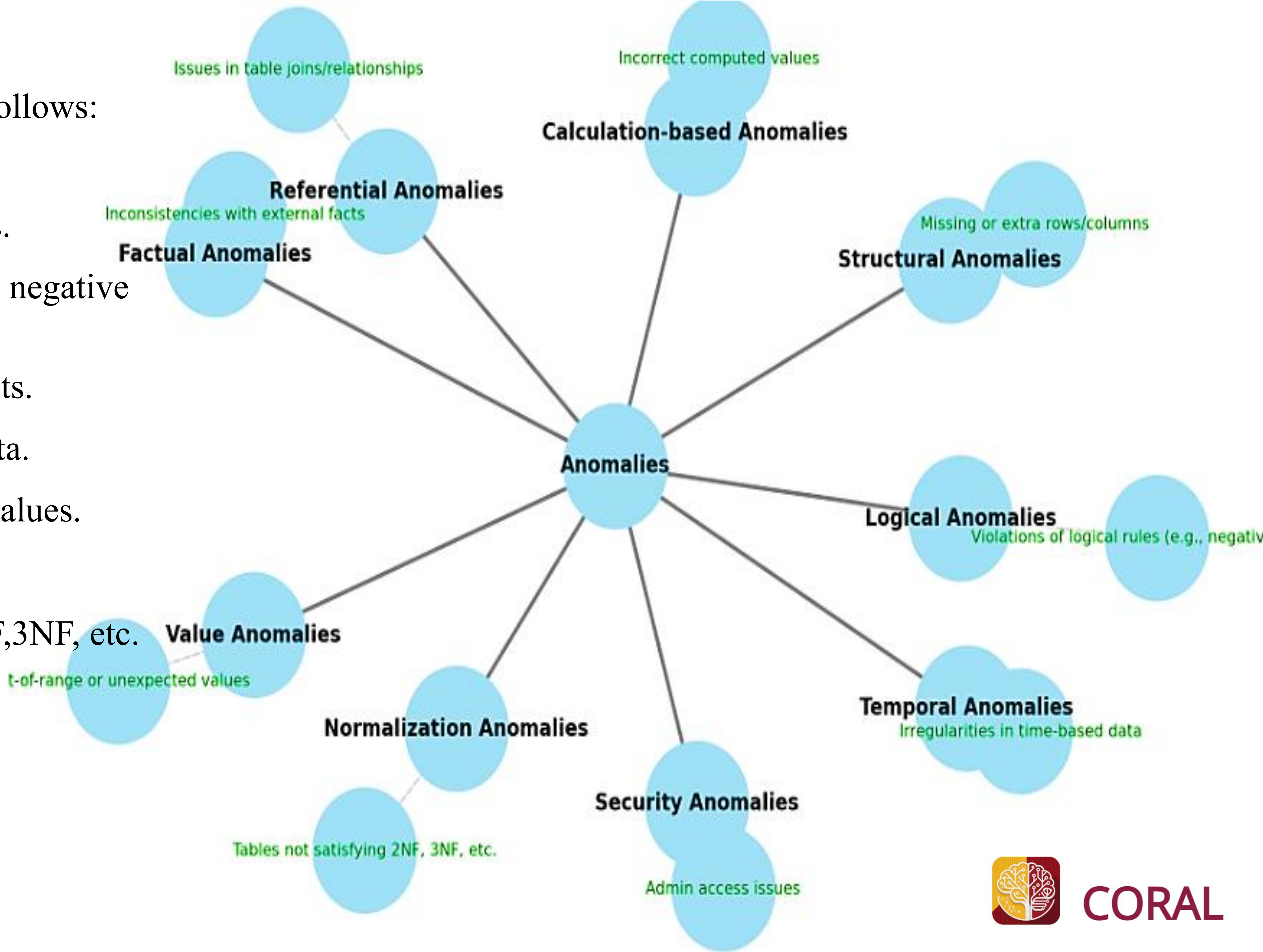
SCAI ASU Ira A. Fulton Schools of Engineering Arizona State University
EMNLP 2025 Suzhou, China 中国苏州
CORAL

Categorization

EMNLP 2025 Suzhou, China | 中国苏州

The anomalies that are covered in this research are as follows:

- **Value Anomalies**: Out-of-range or unexpected values.
- **Logical Anomalies**: Data violating logical rules (e.g., negative salaries).
- **Factual Anomalies**: Inconsistencies with external facts.
- **Temporal Anomalies**: Irregularities in time-based data.
- **Calculation-based Anomalies**: Incorrect computed values.
- **Security Anomalies**: Admin access issues.
- **Normalization Anomalies**: Tables not satisfying 2NF,3NF, etc. forms.



SCAI
ASU Ira A. Fulton Schools of Engineering
Arizona State University

CORAL

Overlapping Cases

Table: Location Data

Location ID	Latitude	Longitude	Description
1	37.7749	-122.4194	San Francisco
2	-95.0000	200.0000	Invalid Location
3	40.7128	-74.0060	New York City

Anomaly:

- **Location ID 2** has an invalid longitude (**200.0000**) since valid longitude values range from **-180 to 180**.

Value Anomaly

Logical Anomaly

Factual Anomaly

Modelling Approaches

Level 1 & 2

Just "Problem" Mentioned

(w/ and w/o CoT): There may be some problems present in the table, without mentioning anomalies or examples.

Anomalies Mentioned

(w/ and w/o CoT): This prompt replaces "problems" with the explicit term "anomalies", providing clearer task framing without examples.

Modelling Approaches

Level 3 & 4

<u>“X” Type of Anomaly Mentioned</u>

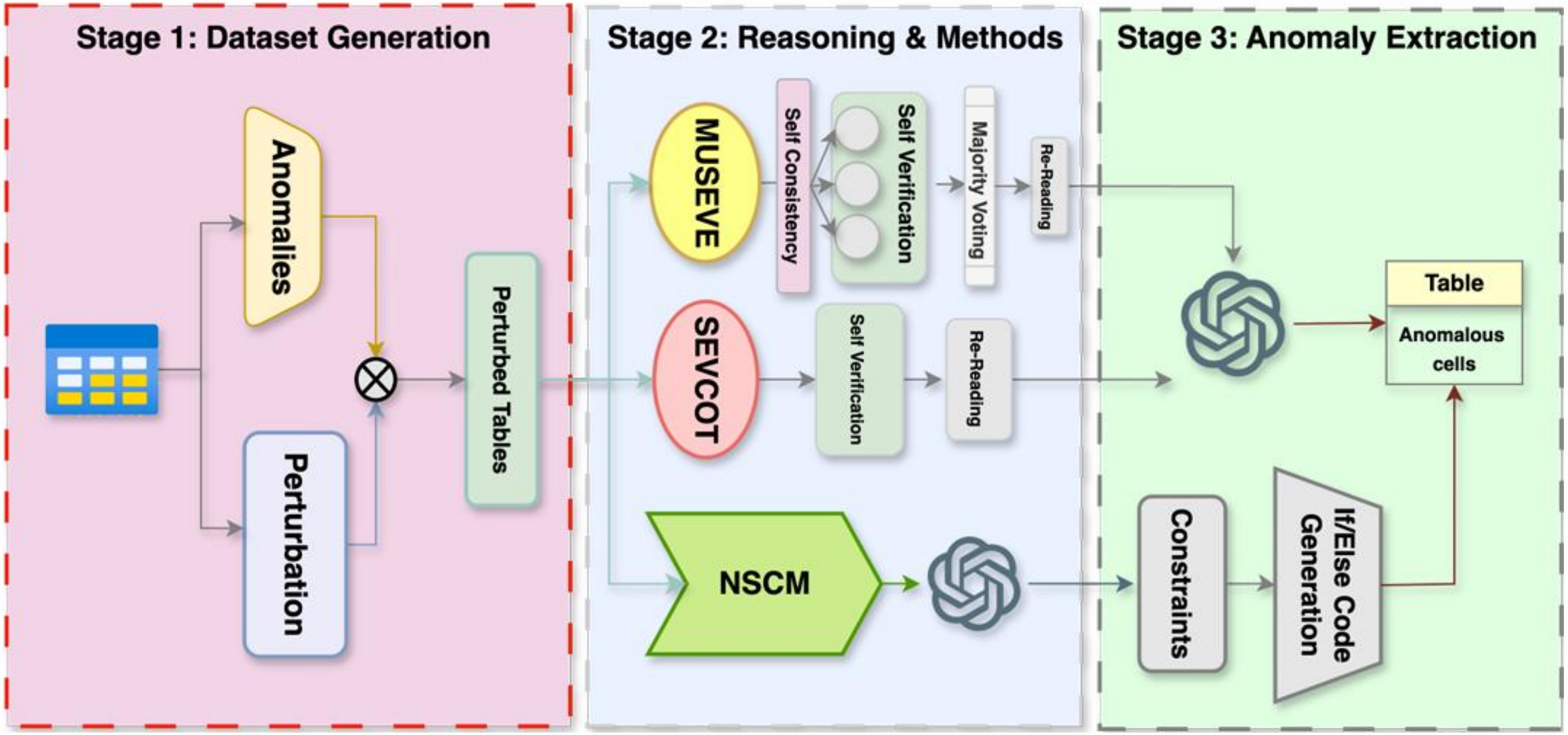
(w/ and w/o CoT): Here, prompt specifies the exact anomaly type (e.g., "factual anomaly", "value anomaly") while still omitting examples.

<u>“X” Type of Anomaly Mentioned with Examples</u>

(w/ and w/o CoT): these prompts enhance specificity further by including both the anomaly type and an illustrative few-shot example

Methods

Diagram



EMNLP 2025 Suzhou, China | 中国苏州

SCAI ASU Ira A. Fulton Schools of Engineering Arizona State University

CORAL

Methods

MUSEVE & SEVCOT

MUSEVE

- Self Consistent prompting with CoT to detect anomalies with different distinct reasoning chains.
- Self verifying the anomalies detected.
- Majority-voting based selection.
- Re-Reading

SEVCOT

- CoT based anomaly detection at granular level.
- Self verifying the anomalies detected.
- Re-Reading

Methods

Neuro Symbolic Constraint Method

LLM + symbolic rules = efficient, interpretable anomaly detection.

Process:

I. From schema S and samples U, LLM generates constraint set:

$$V = \{\varphi_1, \varphi_2, \ldots, \varphi_k\}$$

II. Each φ_i is translated into executable code and run over table D.

III. A cell (r', j) is anomalous if

$$A = \{(r, j) \mid \exists\, \varphi_i \in V,\ \neg\varphi_i(r)\}$$

Example rule:

if *order_date* > *ship_date* ⇒ *flag anomaly.*

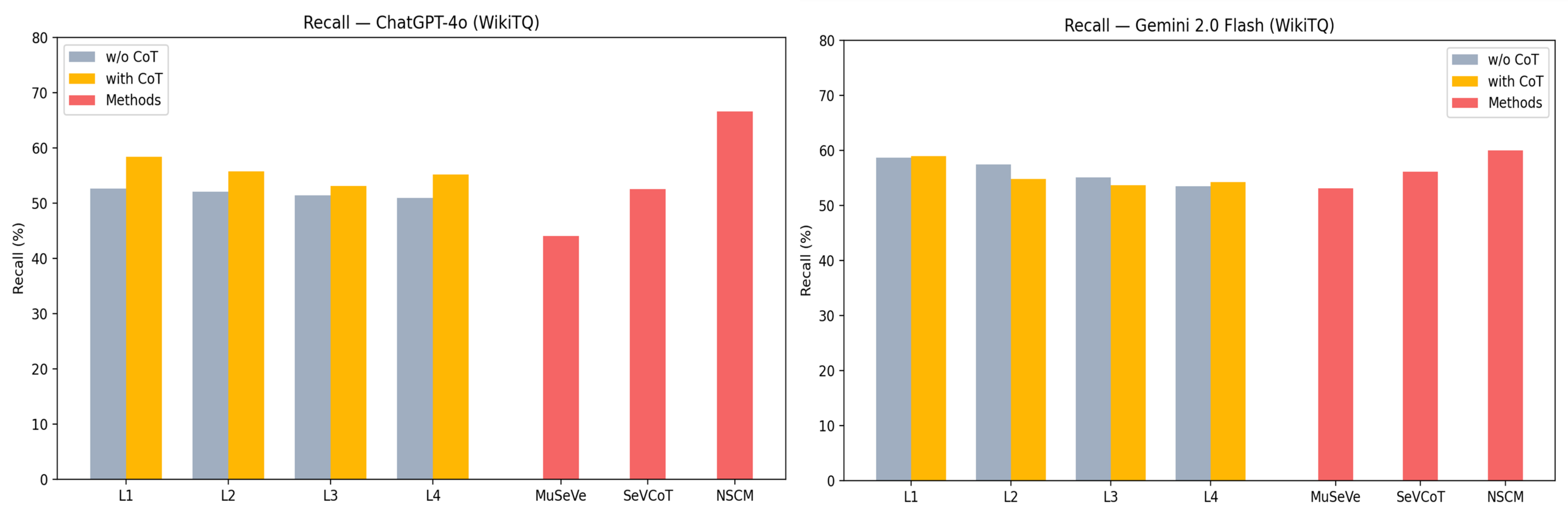
SCAI Ira A. Fulton Schools of Engineering Arizona State University

EMNLP 2025 Suzhou, China | 中国苏州

CORAL

Results

EMNLP 2025 Suzhou, China | 中国苏州



The figures highlights average Recall across various anomaly categories on the WikiTQ dataset, evaluated using four LLMs under different prompting strategies. Li denotes the ith prompt level, with-w/ocot and-wcot indicating absence and presence of Chain-of-Thought reasoning, respectively. MUSEVE and SEVCOT represent multi-reasoning and self-verification variants.

.

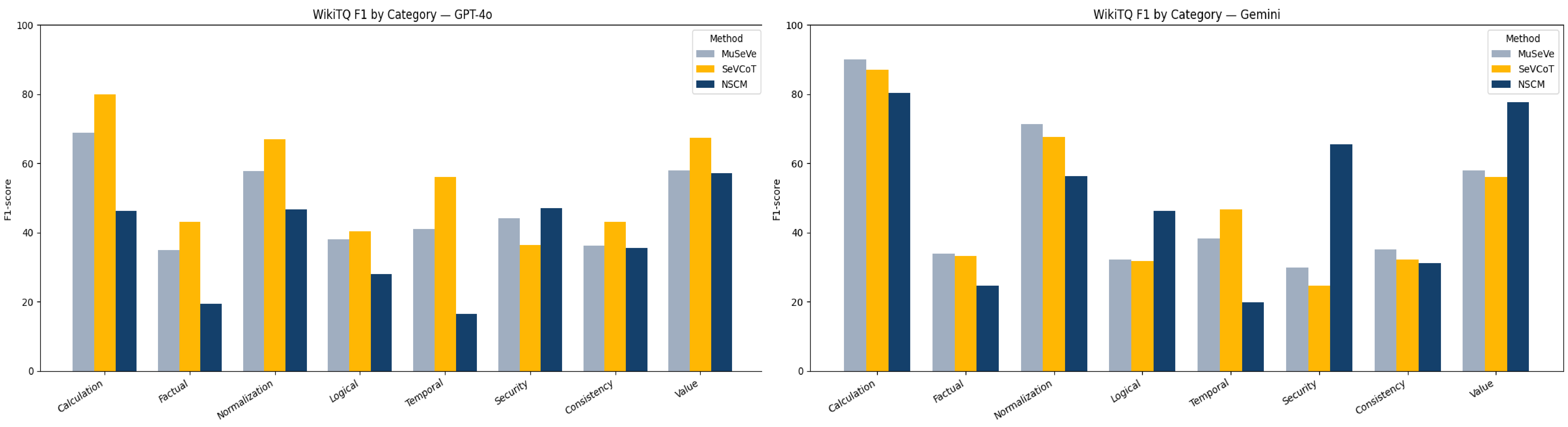
SCAI Ira A. Fulton Schools of Engineering Arizona State University

EMNLP 2025 Suzhou, China | 中国苏州

CORAL

Results

EMNLP 2025
Suzhou, China | 中国苏州



This figure highlights the averaged F1 scores achieved by ChatGPT-4o across eight anomaly categories in the WikiTQ dataset. MUSEVE, SEVCOT, and NSCM represent multi reasoning, self-verification variants, and neuro-symbolic constraint-based methods respectively.

Intrigued? Dive Deeper!

Scan for the Paper at

SCAI

ASU Ira A. Fulton Schools of Engineering

Arizona State University

EMNLP 2025

Suzhou, China | 中国苏州

CORAL

Thank You !

If you have any questions, please feel free to contact me a: **mroycho1@asu.edu**

SCAI
Ira A. Fulton Schools of Engineering
Arizona State University

EMNLP 2025
Suzhou, China | 中国苏州

CORAL