WIP: From Detection to Explanation: Using LLMs for Adversarial Scenario Analysis in Vehicles

David Fernandez, Pedram MohajerAnsari, Amir Salarpour, Cigdem Kokenoz, Bing Li, Mert D. Pesé
Clemson University

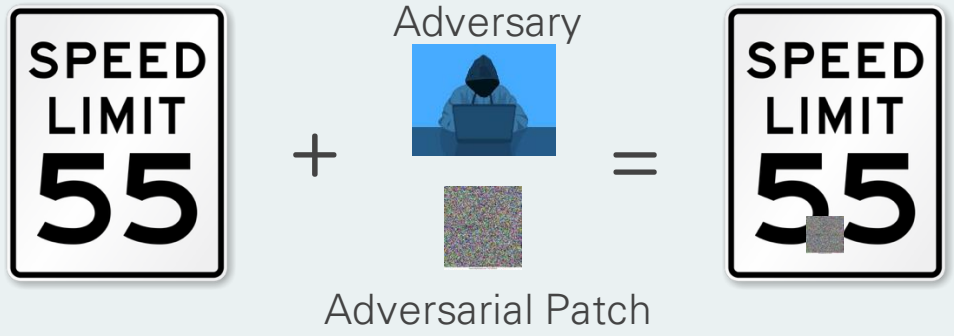
CLEMSON UNIVERSITY

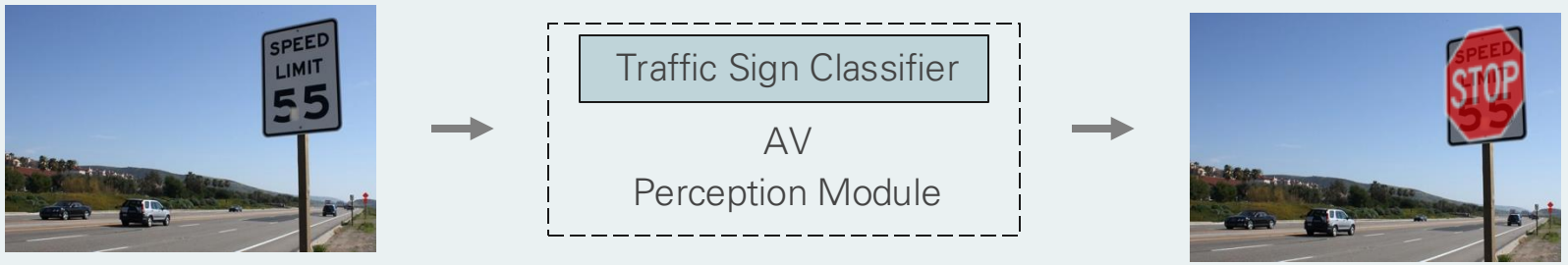
VehicleSec '25

TIGERSEC

Motivation

- DNNs lack robustness against physical-world adversarial perception attacks





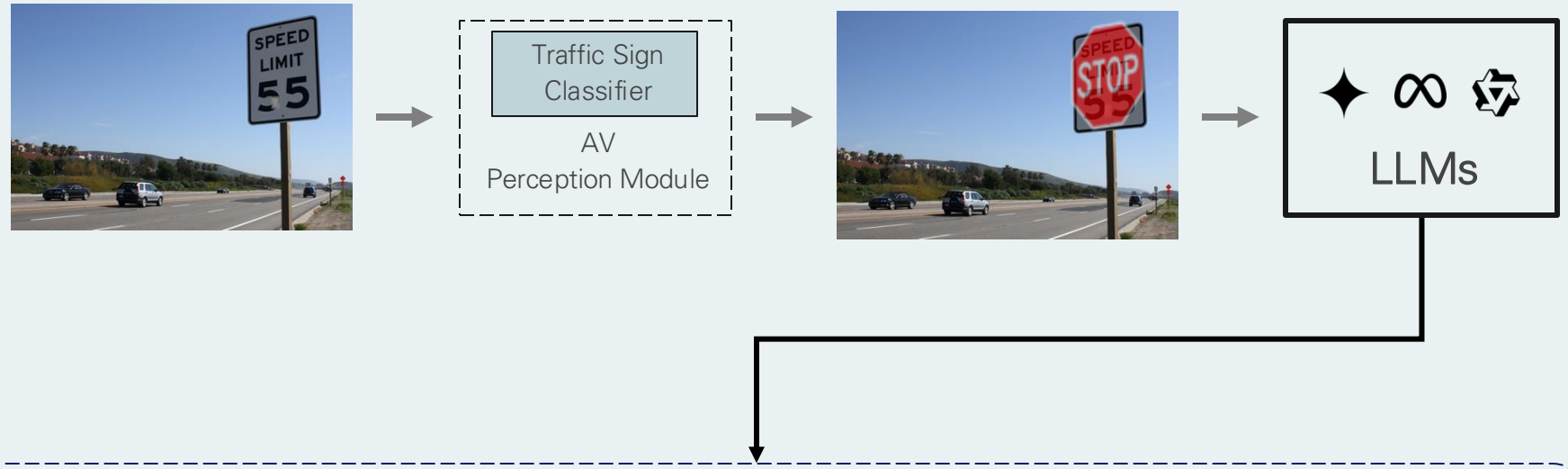
Misclassified traffic sign

Related Work

- Cybersecurity efforts for AVs often focus on anomaly detection, relying on ML for statistical outliers
  - Struggle to detect semantic inconsistencies
  - Require labeled data and fail on unseen attacks
  - Cannot differentiate between or adversarial anomalies
- Recent work shows that LLMs:
  - Can detect anomalies via human-like reasoning
  - Support zero-shot reasoning in time-series and industrial domains

Goal

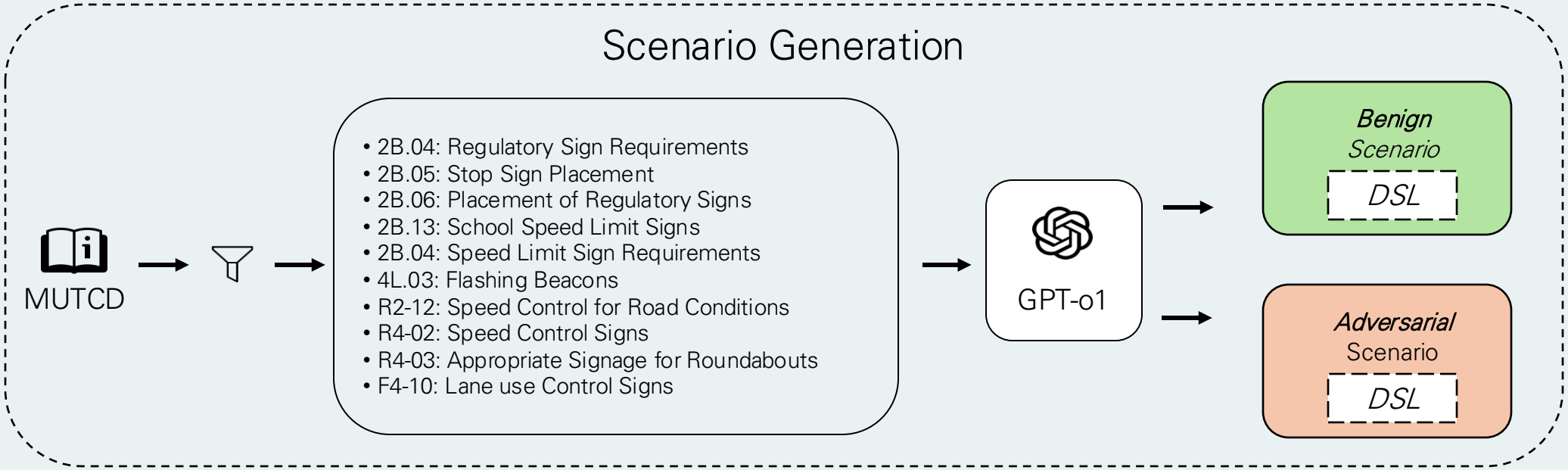
- Framework framework that leverages LLMs for adversarial scenario analysis



*"**Inconsistency Detected:** Placing a stop sign on the roadway shoulder of a highway, with broken white lane markings is inconsistent with standard traffic regulations."*

Dataset Design

- Autonomous Vehicle Security & Explanation Dataset - **AutoSec-X**
  - 40 structured driving scenarios grounded in the MUTCD[1]



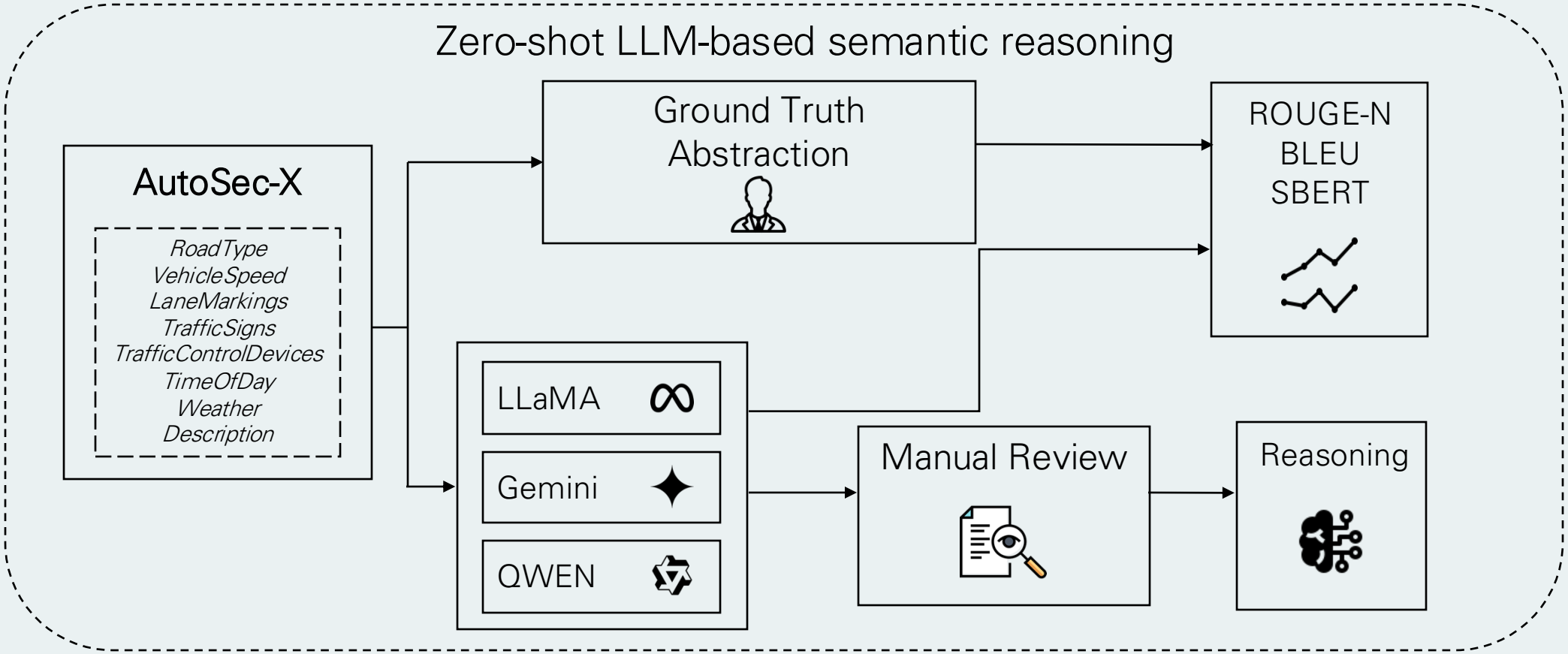
1. Federal Highway Administration. Manual on uniform traffic control devices (mutcd). https://mutcd.fhwa.dot.gov/, 2025. Accessed: 2025-02-15

Scenario Example

- MUTCD Section 2B.05: Standards and guidance for the application of STOP signs.
  - Scenario: *A "Stop" sign appears on a high-speed highway …*

```
{
  RoadType: Highway,
  VehicleSpeed: 65 mph,
  LaneMarkings: Broken white,
  TrafficSigns: [Stop],
  TrafficControlDevices: [],
  TimeOfDay: Day,
  Weather: Clear,
  Description:  Car encounters stop sign in highway
  …
}
```

Scenario Analysis Framework



Ground Truth vs LLM output

Ground Truth

Inconsistency Detected: Yes.

Placing a stop sign on the roadway shoulder of a highway where vehicles can travel at 65 mph speed, with broken white lane markings is inconsistent with standard traffic regulations. According to the MUTCD Sections 2B.04 (Regulatory Sign Requirements) and 2B.05 (Stop Sign Placement), stop signs are intended for use at intersections, ramps, or specific access points where controlled stops are necessary.

Gemini-1.5-flash

Inconsistency Detected: Yes ✓

Reasoning: A stop sign (R1-1) mandates a complete stop, which is inconsistent in a highway. Stop signs are typically placed at intersections or where hazards require a complete stop (MUTCD 2B.04), not on highways designed for high-speed travel. The presence of a stop sign on a highway at 65 mph is a significant traffic regulation violation.

LlaMA-3.2-8B-Instruct

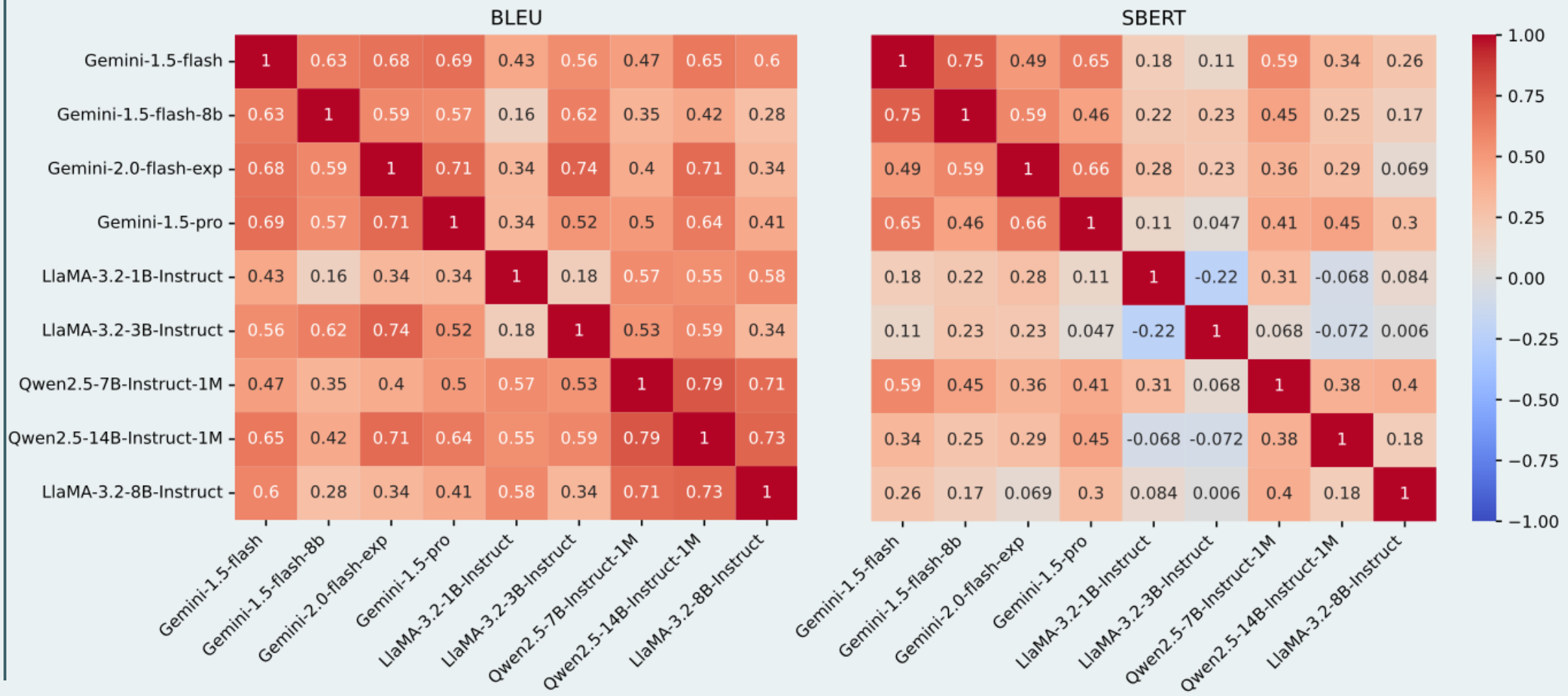
```python def analyze_driving_scenario(scenario):
A stop sign is typically placed at the end of a street or at an intersection, not on the roadway shoulder, which poses safety concerns for drivers and increases the risk of accidents. This placement conflicts with typical real-world traffic laws and guidelines for traffic signs, ✓
indicating an inconsistency in the scenario…

✓ Structured output

✓ Rule relationship

✗ Unstructured output

✗ Hallucinations
```

Accuracy and Semantic Scores

- Gemini models outperform across all metrics
- Natural variance in language generation

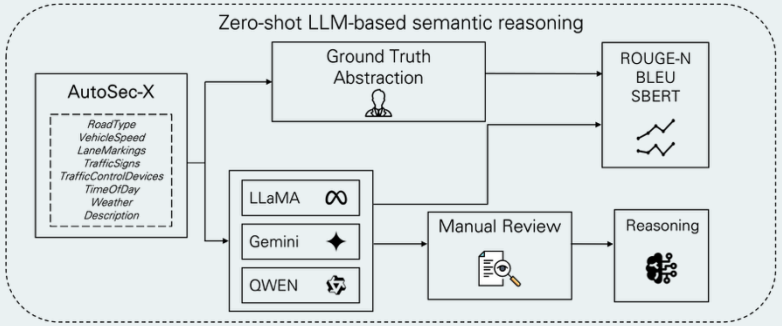
Models	Accuracy	SBERT	BLEU
Gemini-1.5-flash	92.5%	0.680	0.0033
Gemini-1.5-pro	87.5%	0.691	0.003
LLaMA-3.2-8B	82.5%	0.477	0.0008
Qwen-14B	82.5%	0.280	0.0006
LLaMA-3.2-1B	55.0%	0.618	0.0015

Similarity Between Models



Contributions

- Zero-shot CoT LLM-based semantic reasoning approach
- Semantically rich dataset called AutoSec-X
- Robust evaluation strategy that quantitatively and qualitatively compares the reasoning processes of different LLM
- Work bridges the gap between anomaly detection and semantic explanation in AV cybersecurity, with forensic and regulatory applications standing capabilities



{
RoadType: Highway,
VehicleSpeed: 65 mph,
LaneMarkings: Broken white,
TrafficSigns: [Stop],
TrafficControlDevices: [],
TimeOfDay: Day,
Weather: Clear,
Description: Car encounters stop sign in highway
}

Future Work

- Expand the dataset to include scenarios from existing datasets

- Incorporate Scenic to generate visualizations from DSL

- Explore VLM scene interpretation as an anomaly forensic analysis tool.

David Fernandez
dferna3@clemson.edu

Pedram MohajerAnsari
pmohaje@clemson.edu

Dr. Mert Pese
mpese@clemson.edu

Thank you!

CLEMSON UNIVERSITY

VehicleSec '25

TIGERSEC