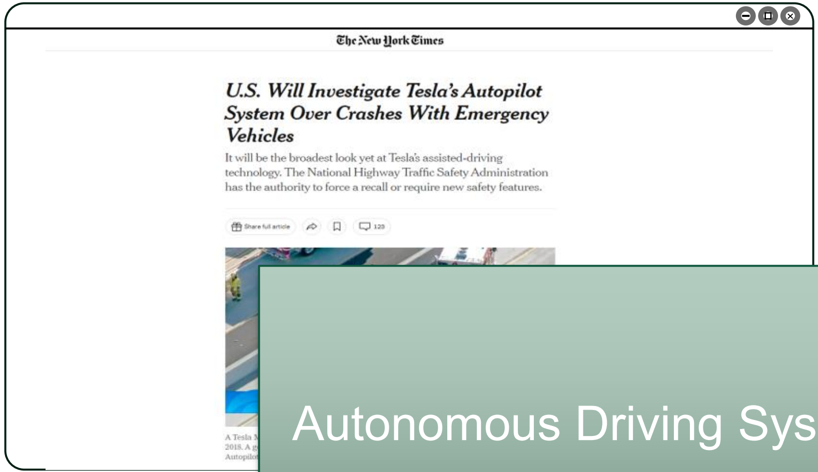
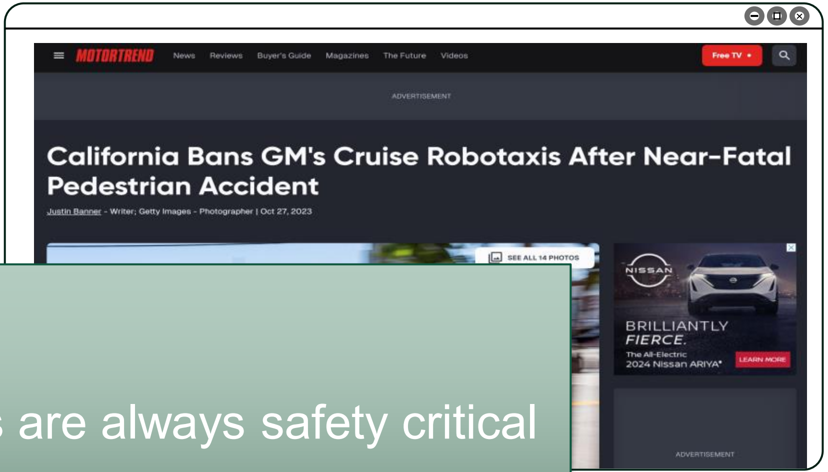
LLMs for Autonomous Driving Systems

Trey Woodlief

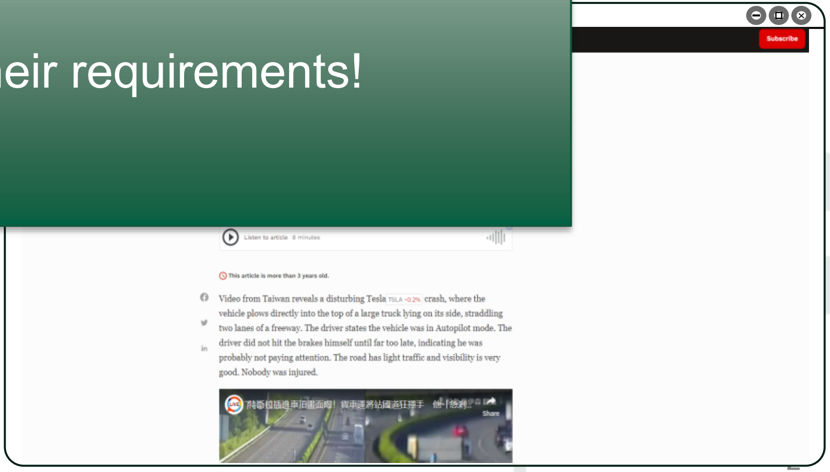
WILLIAM & MARY
CHARTERED 1693





Autonomous Driving Systems are always safety critical

They need to meet their requirements!



How do we leverage requirements?

- Requirements allow us to:
  - Design & build
  - Test
  - Monitor

Improving ADS Safety

Requirements are about the ADS in the world!

safety

are met

in the physical world

What's in a requirement?

§ 46.2-821 – The driver of a vehicle approaching an intersection on a highway controlled by a stop sign shall, immediately before entering such intersection, stop

Formalization in temporal logic:

$$\mathrm{G}((\neg\, \boldsymbol{hasStop} \land \mathrm{X}\, \boldsymbol{hasStop}) \rightarrow (\mathrm{X}\, \boldsymbol{hasStop}\ \mathrm{U}\ (\boldsymbol{isStopped} \lor \mathrm{G}\, \boldsymbol{hasStop}))$$

What's in a requirement?

§ 46.2-821 – The driver of a vehicle approaching an intersection on a highway controlled by a stop sign shall, immediately before entering such intersection, stop

VIRGINIA LAW

Formalization in temporal logic:

$$G((\neg\, \mathit{hasStop} \land X\, \mathit{hasStop}) \rightarrow (X\, \mathit{hasStop}\ U\ (\mathit{isStopped} \lor G\, \mathit{hasStop})))$$

What's in a requirement?

VIRGINIA LAW

§ 46.2-821 – The driver of a vehicle approaching an

shall,

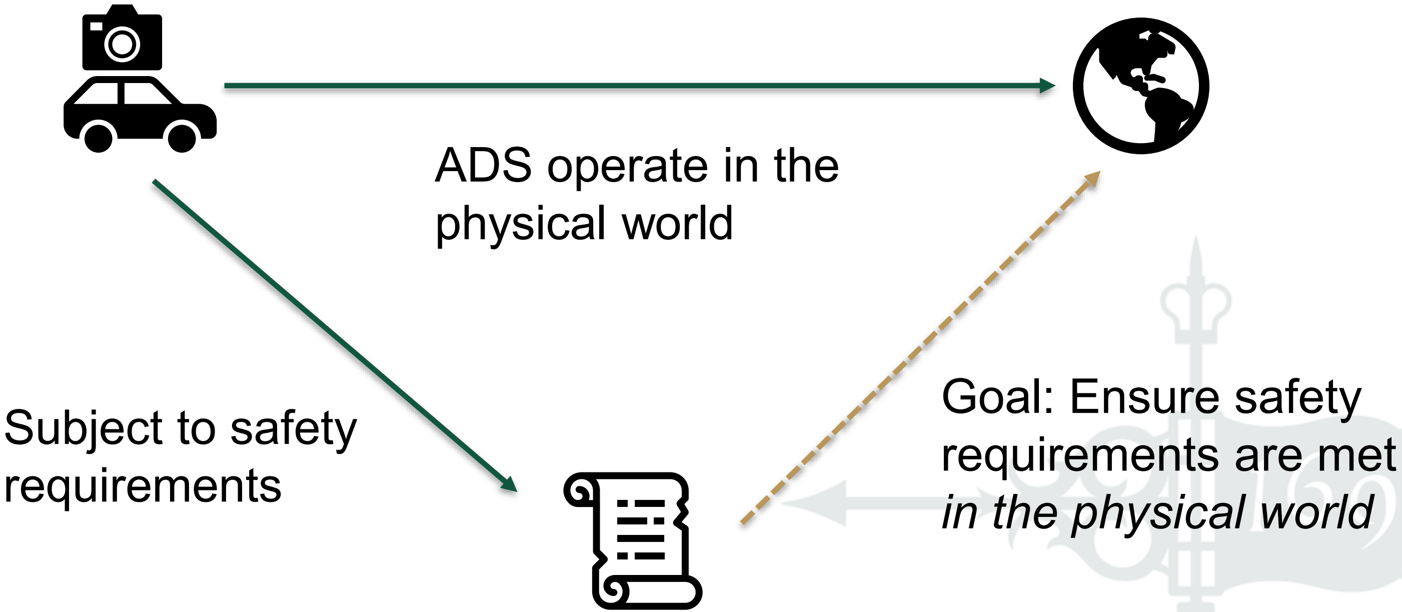
Temporal formalization is precise,
but still relies on the physical interpretation!

Formalization

G((¬ *hasSt*

→ (X *hasSt*

Improving ADS Safety



ADS Requirements

- Always described relative to the world
- In terms originally designed for humans
  - *Stop at the stop sign*
  - *Do not drive in heavy rain*
  - *Do not get too close to pedestrians*

ADS Requirements

- Always described relative to the world
- In terms originally **designed for humans**
  - *St*
  - *Do*
  - *Do*

LLMs are designed to mimic humans

Can we leverage LLMs to use ADS requirements?

How do LLMs work?

- LLMs are prediction engines
- LLMs learn by example
  - Learn *embeddings*
  - Learn *associations*

What do LLMs learn?

What do LLMs learn?

Embeddings

Goal: represent words so

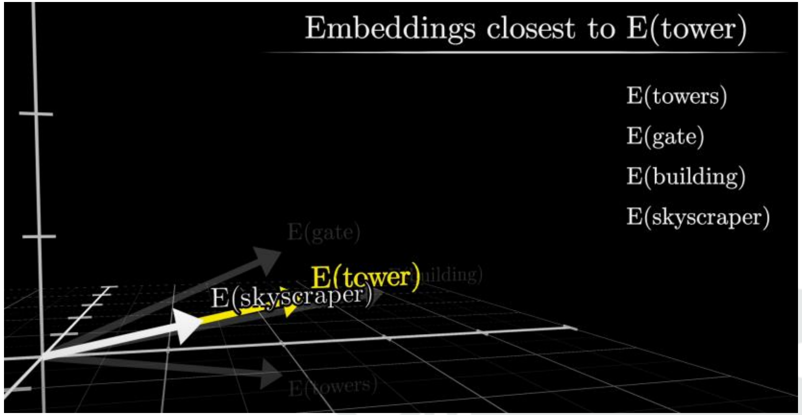
1. Similar concepts are near each other
2. The embedding is also a *vector* carrying its meaning

What do LLMs learn?

Embeddings

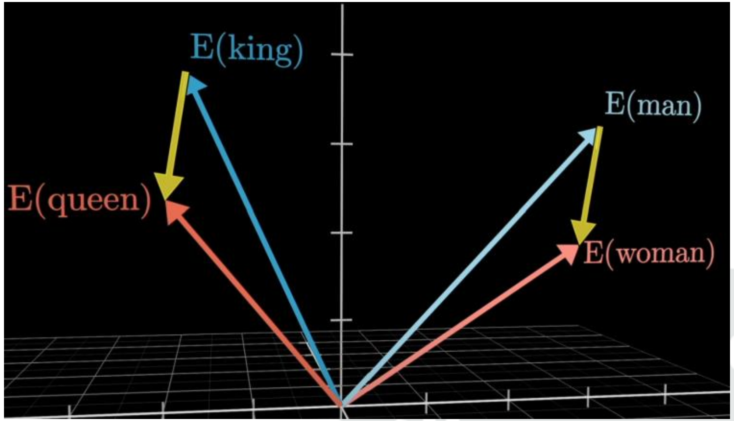
Internet

LLM



What do LLMs learn?

Embeddings



What do LLMs learn?

Associations

Goal: what is the most likely next word?

What do LLMs learn?

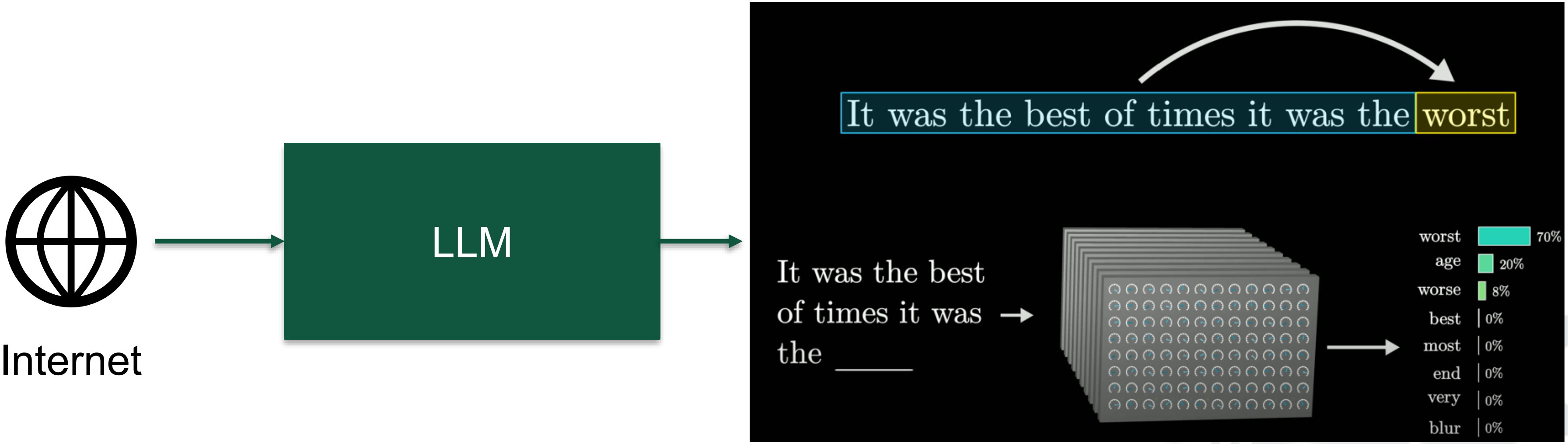
Associations

Goal: what is the most *likely* next word?

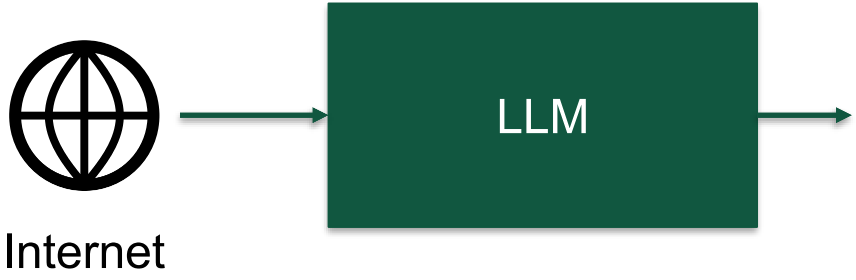
What defines likely?
Prior human examples!

What do LLMs learn?

Associations



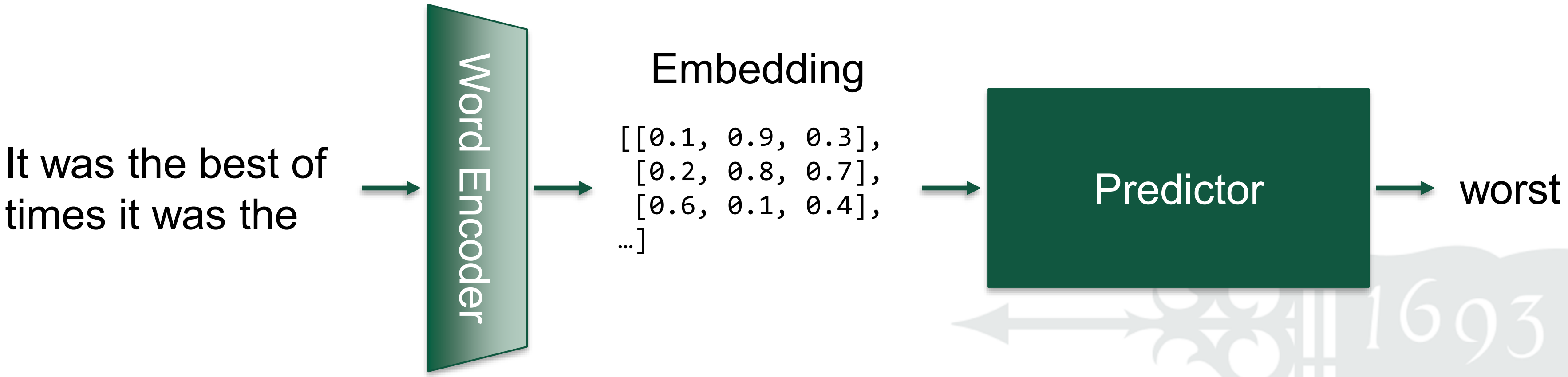
What do LLMs learn?



By training on prior human examples, LLMs learn:

1. Semantics of words relative to each other
2. To predict the next word from context

How do LLMs work?



The Visual Variant

- Vision Language Models
  - Learn to interpret images by embedding

The Visual Variant

What sign is this?

Word Encoder

Image Encoder

Embedding

```
[[0.3, 0.4, 0.3],
 [0.2, 0.3, 0.7],
 [0.1, 0.1, 0.4],
…]
```

```
[[0.5, 0.2, 0.3],
 [0.2, 0.8, 0.5],
 [0.6, 0.1, 0.4],
…]
```

Predictor

stop

Leveraging LLMs for ADS

- Use LLMs to:
  - Interpret the world
  - Directly resolve requirement ambiguity
  - Formalize requirements

ODD-diLLMma

- ODD compliance checking using LLMs

ODD-diLLMma

- ODD compliance checking using LLMs
- Operational Design Domain
  - Defines environment conditions for ADS
  - Requirement for operation

ODDs are defined in **natural language** about the **world**

What is in ODD?

Company	DAS	ODD Semantic Dimension									
		Weather and Environmental Factors						Roadway Characteristics			
		Heavy Rain	Rain	Sleet	Ice	Bright Light	Low Light	Sharp Curves	On-Off Ramps	Intersections	Traffic Signals
comma.ai [9]	A									×	
	A										×
GMC [18]	Supe									$\times^{c}$	×
	Traff Cruis										
	Au										
								×			
T									✓	✓	✓

How do we identify these in the field?

How close to an intersection counts?

arp curve?

How b

ODD-diLLMma

- ODD compliance checking using LLMs
  - LLMs can interpret the world
  - LLMs have seen previous human judgement to decide these ambiguities

Use LLMs to check sensor data against the ODD!

ODD-diLLMma

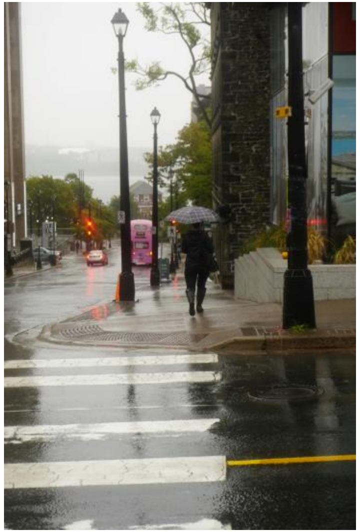
- ODD compliance checking using LLMs

ODD-diLLMma

The ADS cannot
operate in heavy rain.

ODD-diLLMma

- ODD compliance checking using LLMs



ODD-diLLMma

The ADS cannot
operate in heavy rain.

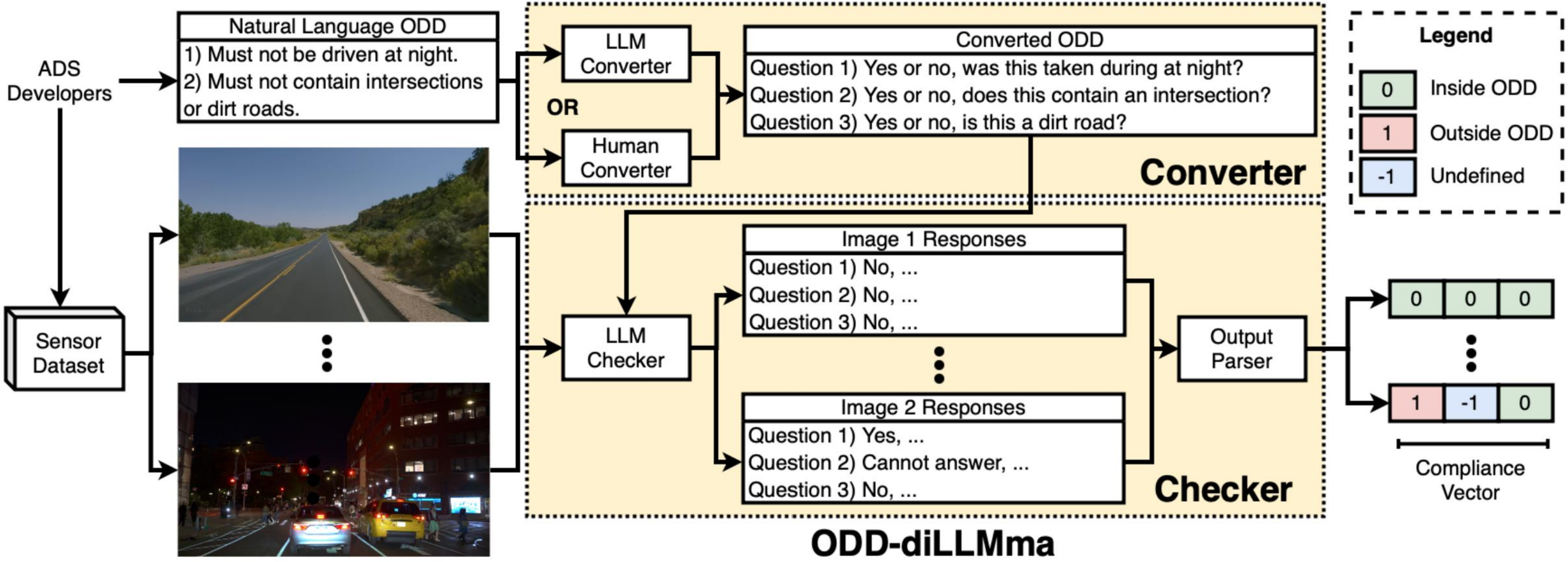
ODD-diLLMma

- ODD compliance checking using LLMs

ODD-diLLMma

The ADS cannot
operate in heavy rain.

ODD-diLLMma


Natural Language ODD
1) Must not be driven at night.
2) Must not contain intersections or dirt roads.
ADS Developers
Sensor Dataset
LLM Converter
OR
Human Converter
Converted ODD
Question 1) Yes or no, was this taken during at night?
Question 2) Yes or no, does this contain an intersection?
Question 3) Yes or no, is this a dirt road?
Converter
Legend
0 Inside ODD
1 Outside ODD
-1 Undefined
LLM Checker
Image 1 Responses
Question 1) No, ...
Question 2) No, ...
Question 3) No, ...
Image 2 Responses
Question 1) Yes, ...
Question 2) Cannot answer, ...
Question 3) No, ...
Output Parser
0 0 0
1 -1 0
Compliance Vector
Checker
ODD-diLLMma

ODD-diLLMma

- ODD compliance checking using LLMs
  - Can we leverage this in practice?

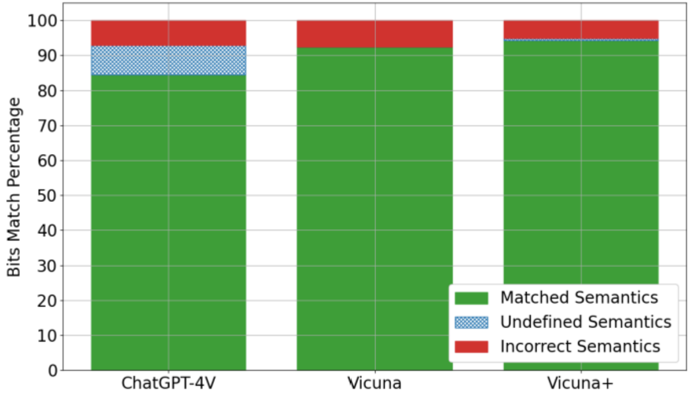


Fig. 8: ODD-diLLMma Semantic Accuracy by LLM

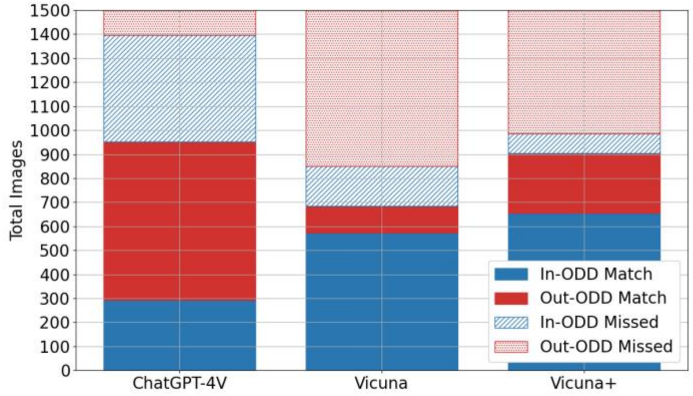


Fig. 7: ODD-diLLMma in-ODD Accuracy by LLM

ODD-diLLMma

- ODD compliance checking using LLMs
  - Can we leverage this in practice? *Carefully*

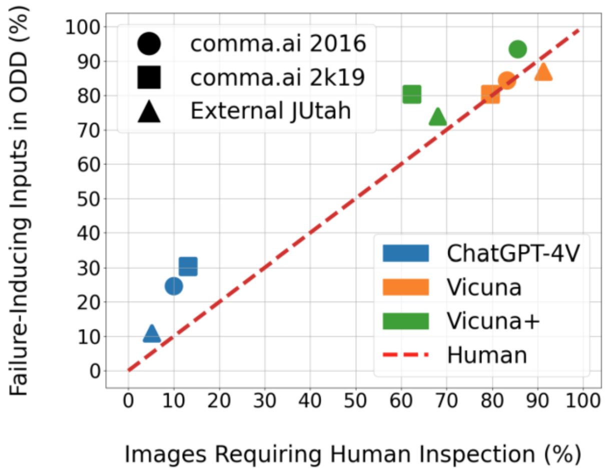


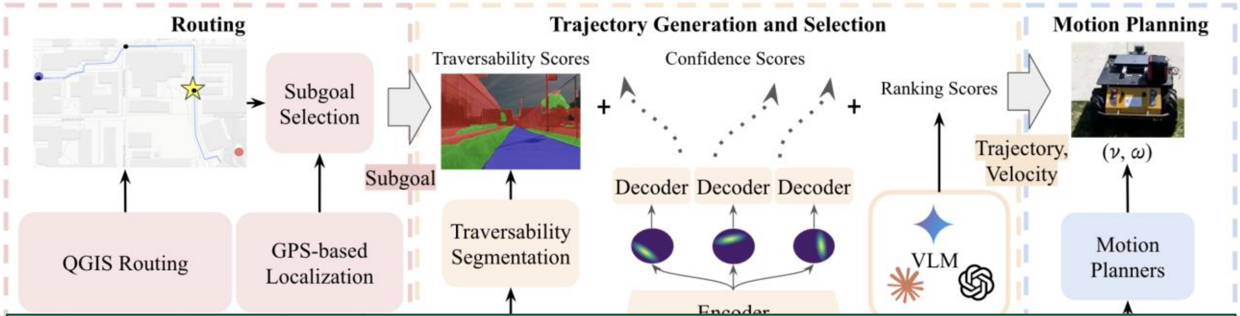
Fig. 6: In-ODD Failures vs Images Requiring Inspection

ODD-diLLMma

- Operationalize Natural Language Reqs.
- Over the real world through sensor data
- Mimicking human labeling/decisions

MOSU

- Multi-modal perception and On-road Scene Understanding for mobile robots



Goal: find a trajectory for the robot that meets the reqs

range waypoints, serving as high-level guidance for its trajectory generation system.

MOSU

- Multi
  Scen
  ts

VLMs have observed this behavior – can they decide the requirement in practice?

- Requirements:
  - Traversable terrain
  - Obeys traffic laws
  - *Obeys social cues*

Personal space is a cultural norm.

Robot must understand social cues and cultural norms for people to be comfortable around the robot.

The N trajectories are labeled with numbers [0-N-1] from right to left in sequence. The goal is K meters at [Right Front]. Rank trajectories for social navigation.

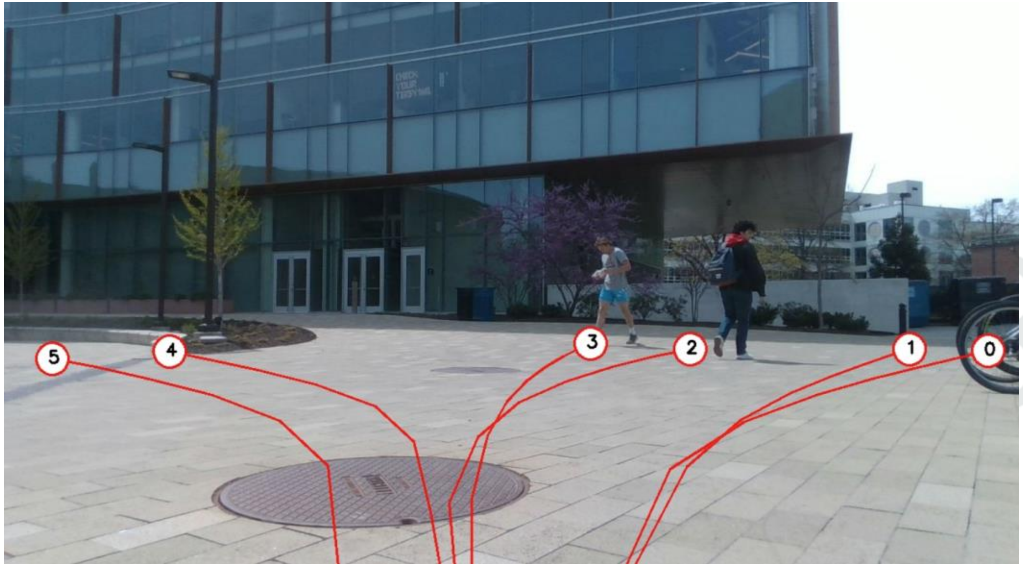
1. keep away from the groups of pedestrians. The robot has three mode, Normal, Slow, and Stop. If the people are approaching, the robot needs to Slow. If people are too close or there is no open space, the robots Stops.
2. follow the traffic rules, and if going across the street, the robot should keep in crosswalks.
3. recognize the traffic signs and behave accordingly.
4. avoid off-road terrain for small wheeled robots.

Given the picture, the target is at K meters Front Left. Rank the trajectories by the criteria. output the format: [robot mode], [ranked numbers], reason

MOSU

- Multi-
  Scen ts

Does the paper defend these rankings?



MOSU

- Operationalize “unwritten rules”
- How would we formalize these?
- What is the baseline for performance?

ScenicNL

- Translating real-world crashes into tests

ScenicNL

- Translating real-world crashes into tests
- Even for formalized requirements, testing is hard
  - Complex environment
  - Many actors

How do we write down test cases anyway?

ScenicNL

- Translati
- Scenic
  - Probabilistic programming language
  - Describes distributions of *scenarios*
  - A test case samples from the distribution
  - Ready-made integration with simulators

How do we write down test cases anyway?

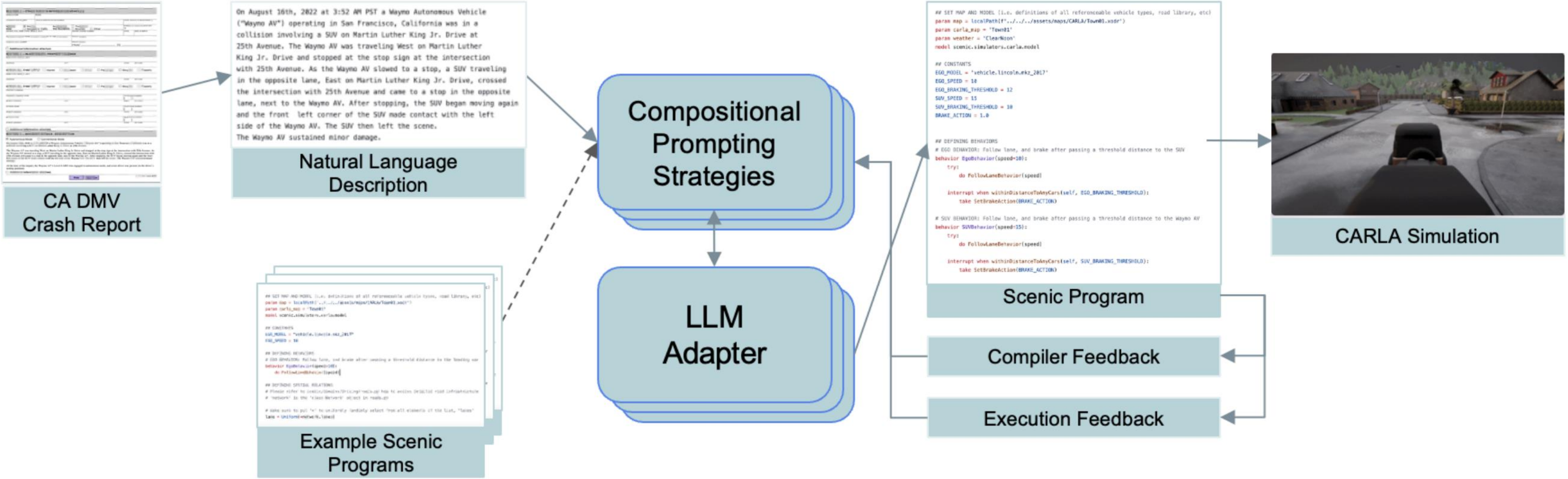
ScenicNL

- Translating real-world crashes into tests
- Humans have been driving for a long time
- Crashes represent difficult scenarios
- DMV has been documenting for years

How do we know what makes a good test case?

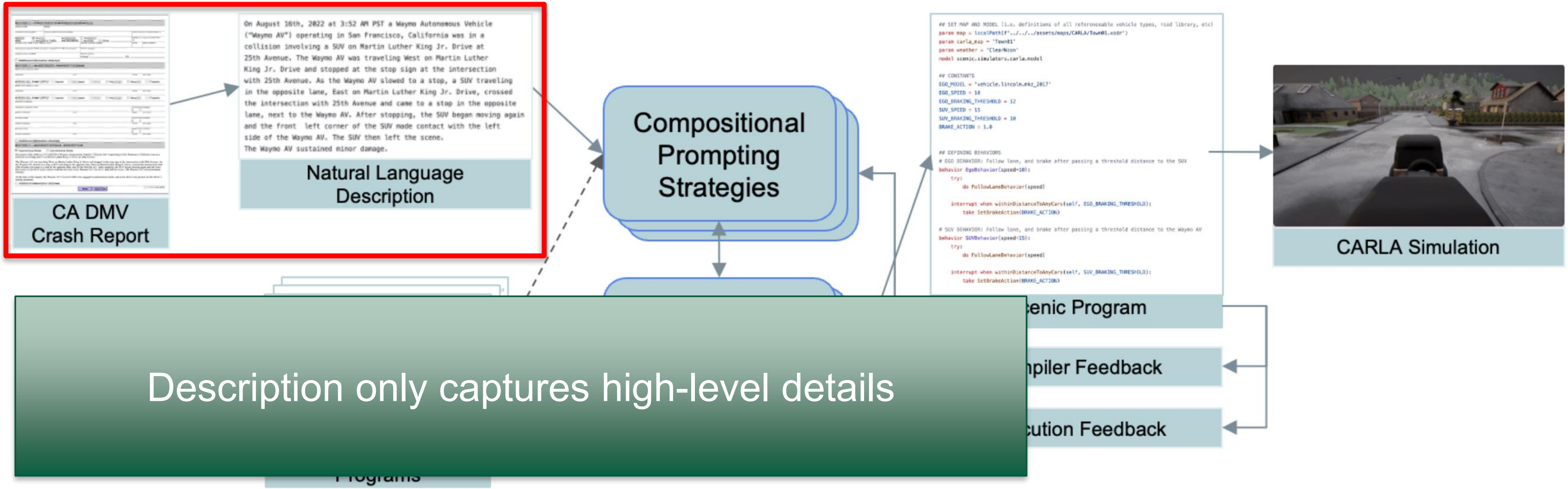
ScenicNL

- Translating real-world crashes into tests



ScenicNL

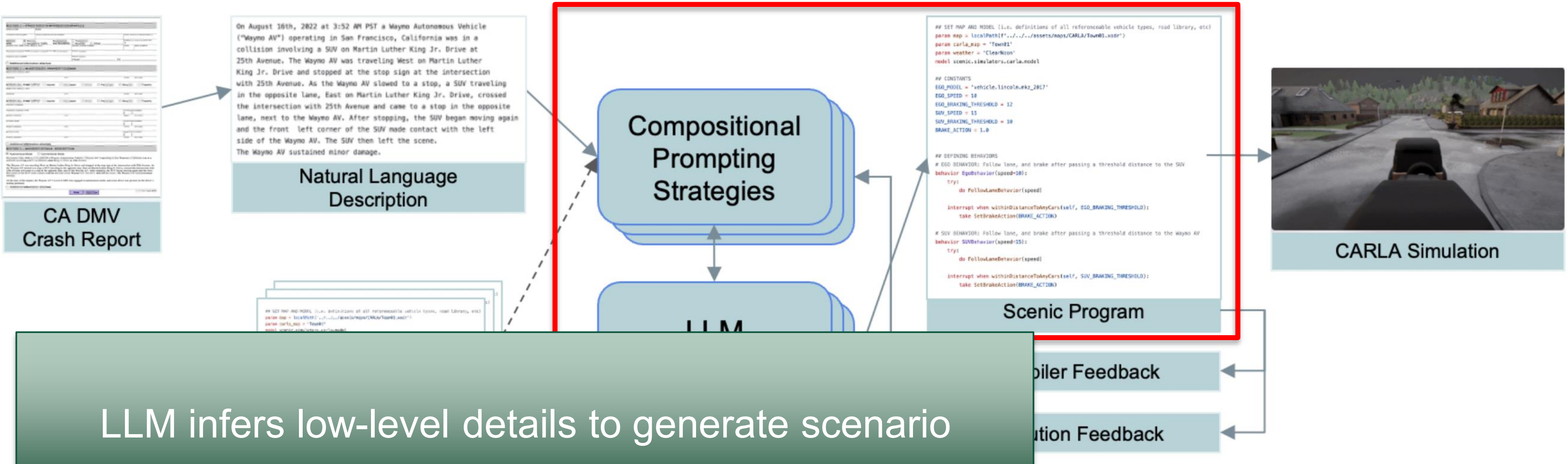
- Translating real-world crashes into tests



Description only captures high-level details

ScenicNL

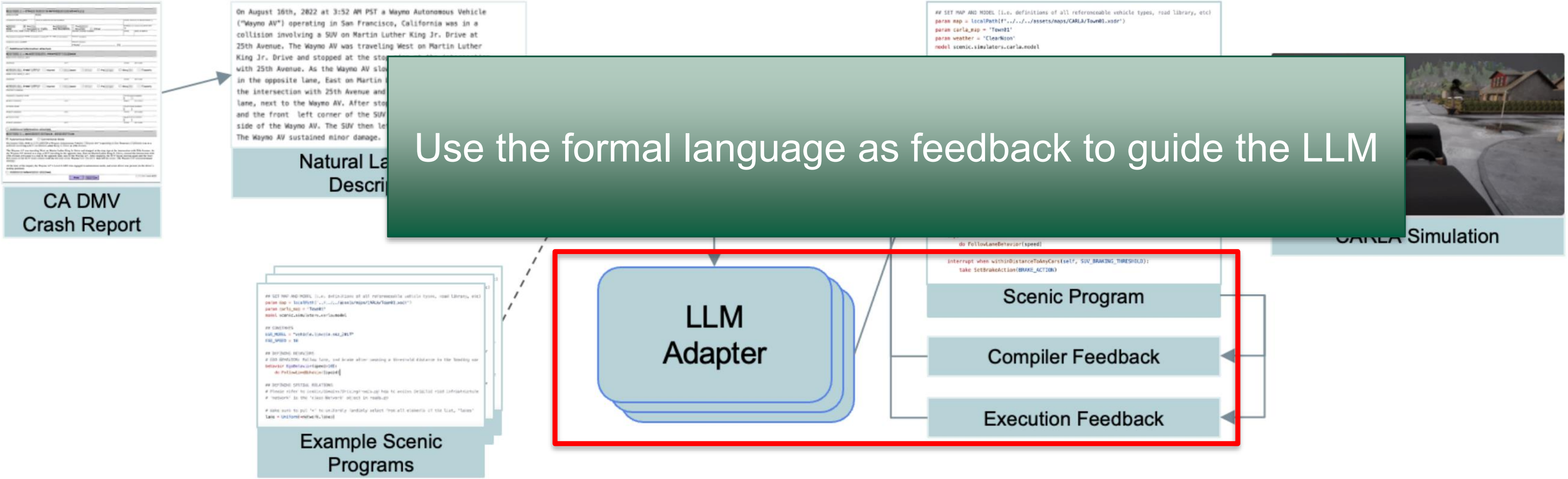
- Translating real-world crashes into tests



LLM infers low-level details to generate scenario

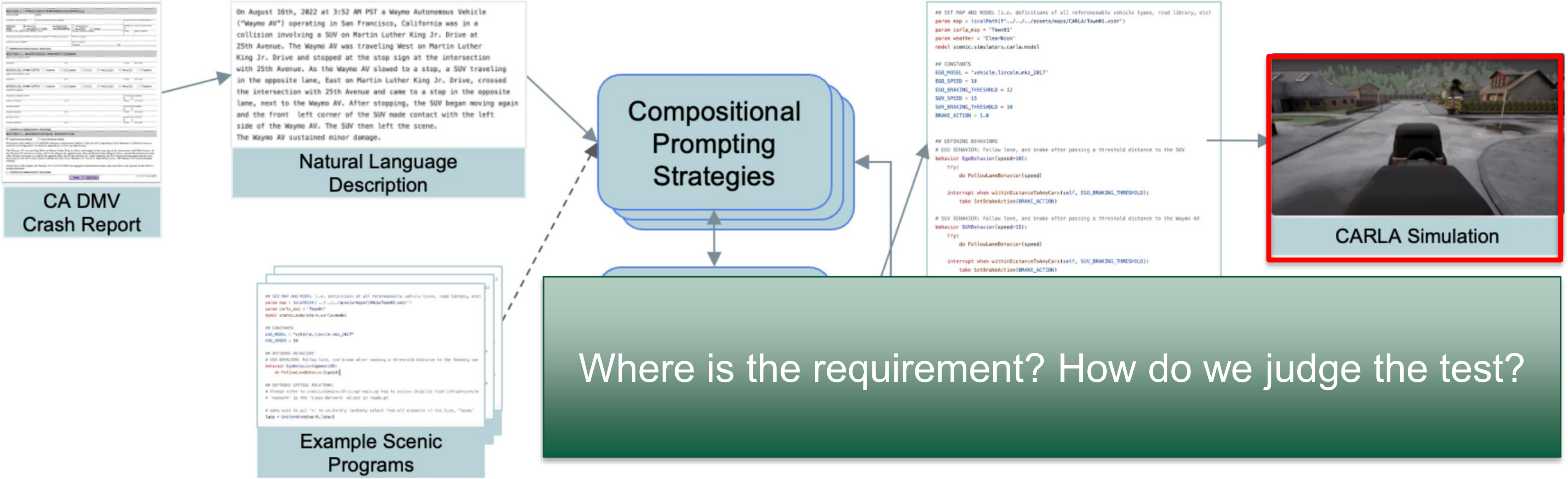
ScenicNL

- Translating real-world crashes into tests



ScenicNL

- Translating real-world crashes into tests



Where is the requirement? How do we judge the test?

ScenicNL

- Leverage prior natural language
- Generate known-difficult test cases
- Automatically test ADS

Where can LLMs help?

- Interpret complex sensor data
- Directly decide ambiguous requirements
- Reason through informal/unwritten reqs
- Translate natural language to formal reqs