

From Crash Reports to Safer Roads:

A Multimodal Framework Integrating Vision-Language Models and Street View Analysis

Guanhe Wu

Graduate Research Assistant
guanhe_wu@student.uml.edu

Xintong Yan

Ph.D.
Postdoctoral Research
Associate
Xintong_yan@uml.edu

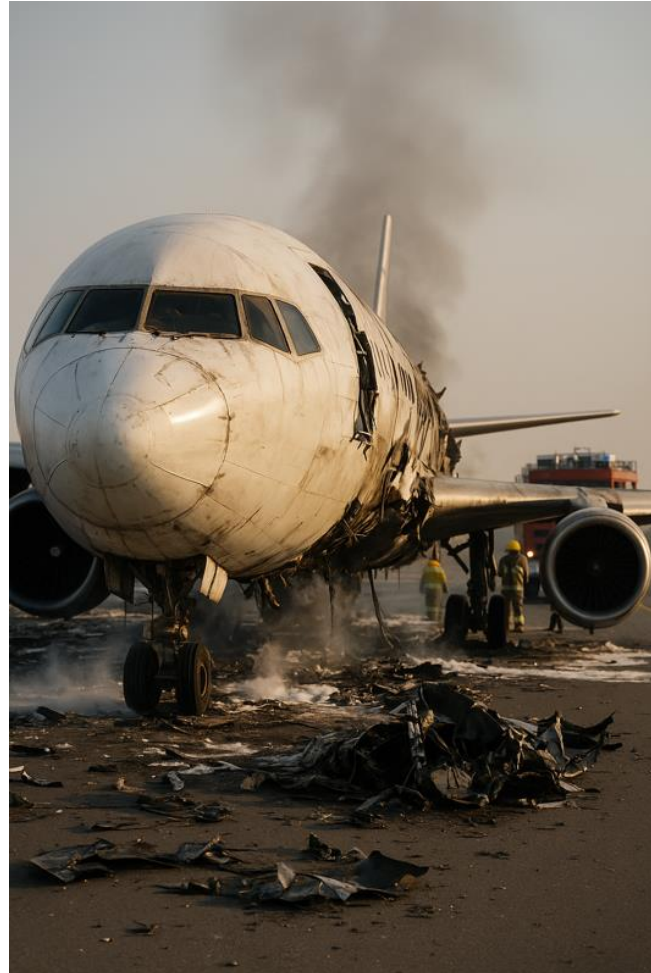
Yuanchang Xie

Ph.D., P.E.
Professor
Yuanchang_xie@uml.edu



Aviation Accidents vs Road Crashes

Aviation Accidents

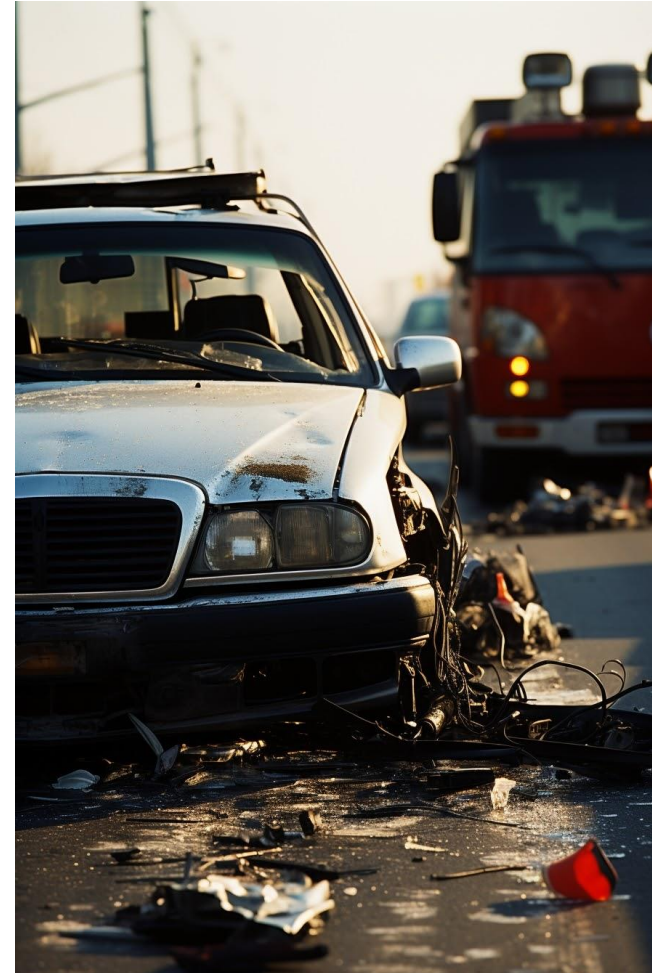


Rare but serious

In-depth investigation

Complete causal chain

Road Crashes



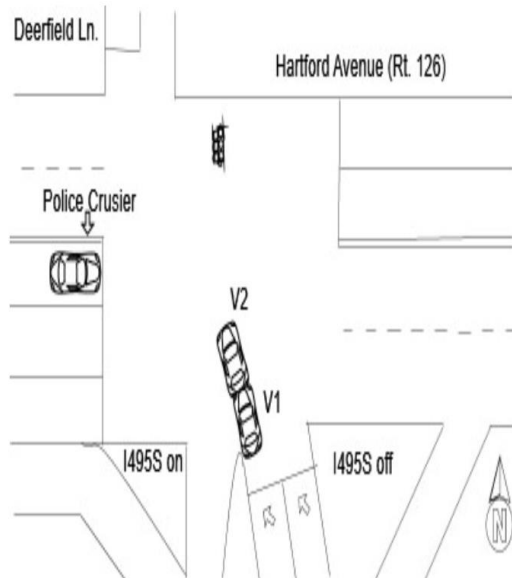
More Frequent

Statistical level

Lack causal detail

Tabular Data vs Crash Reports

CRASH NUMB	CRASH DATE	CRASH_TIME	CRASH_HOUR	CRASH_STATUS	CRASH_SEVERITY	NUMB_VEH	NUMB_NONFATAL_INJR	NUMB_FATAL_INJR	MANR_COLL_DESCR	AMBNT_LIGHT_DESCR	WEATH_COND_DESCR	ROAD_SURF_COND_DESCR	...
5195046	1/1/23	1:15:00	1:00:00	Open	PDO	3	0	0	Rear-end	Dark - lighted roadway	Rain/Cloudy	Wet	...
5196046	1/1/23	1:18:00	1:00:00	Open Fatal	FI	2	1	1	Head-on	Daylight	Clear	Dry	...
5196045	1/1/23	8:12:00	8:00:00	Open Fatal	FI	1	0	1	Single vehicle crash	Dark - lighted roadway	Clear/Clear	Dry	...
5196037	1/1/23	12:23:00	12:00:00	Open	FI	4	1	0	Rear-end	Dark - lighted roadway	Cloudy	Wet	...

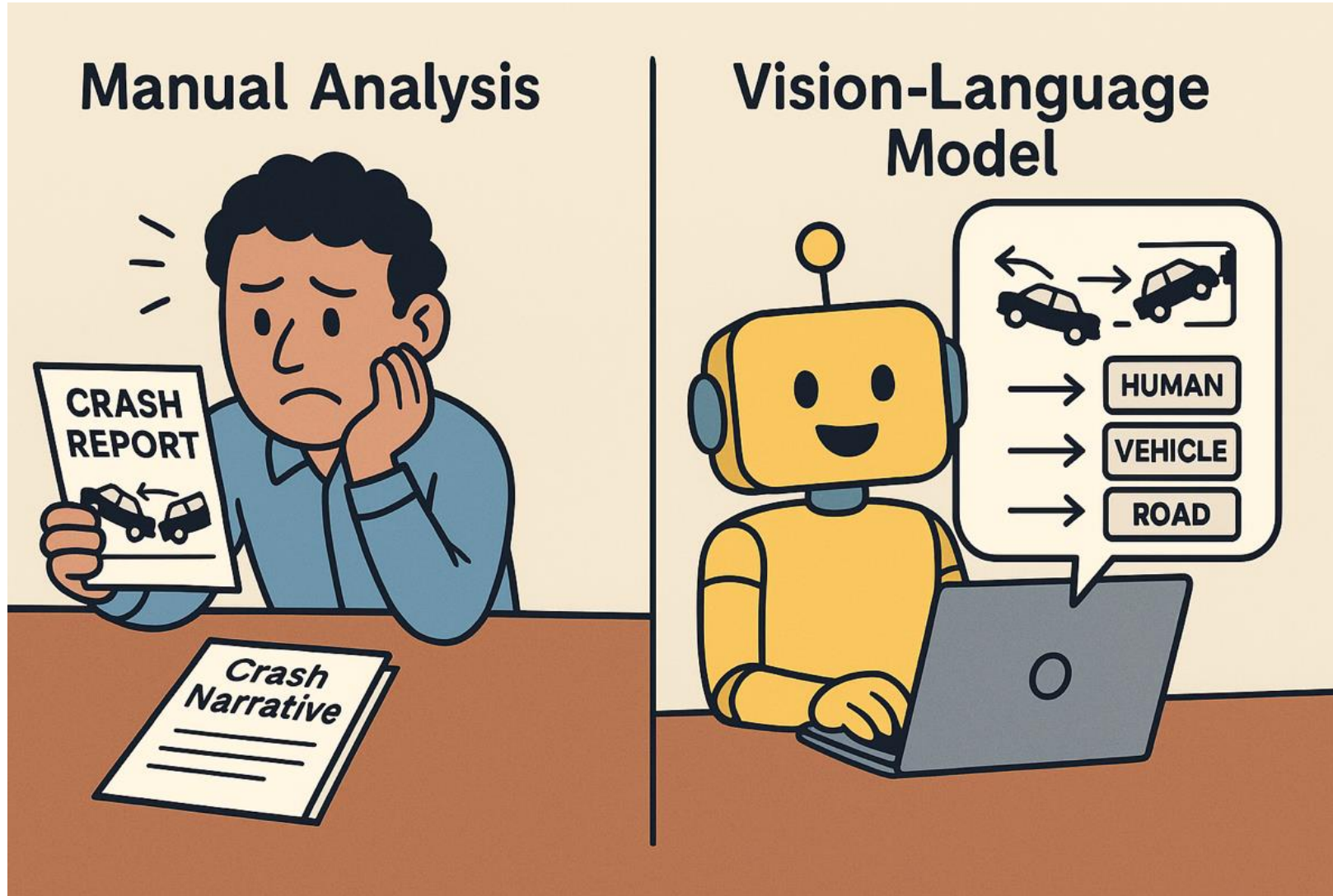


Crash Narrative:

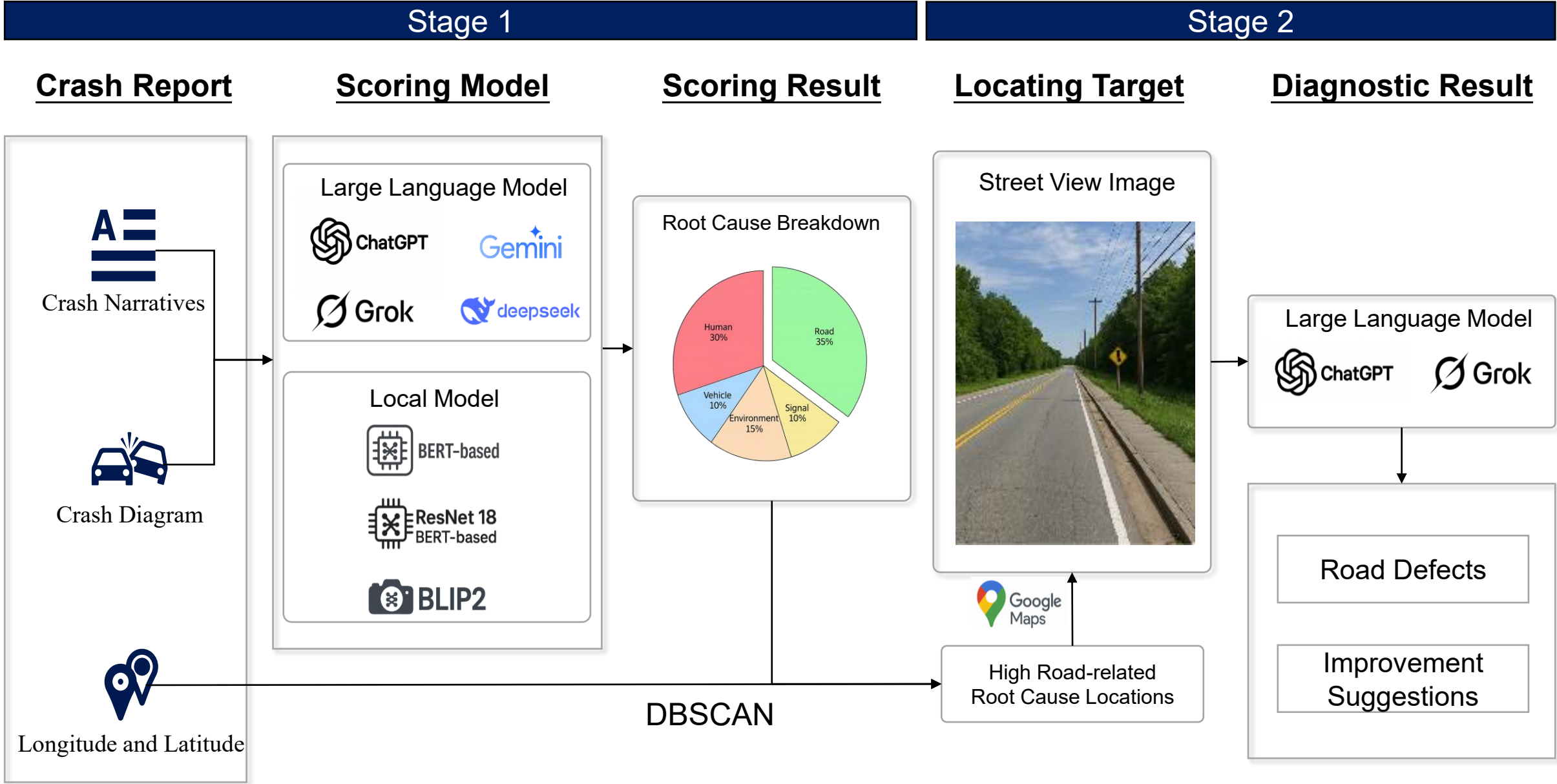
V1 AND V2 WERE COMING OFF THE SOUTHBOUND 495 OFF RAMP, TURNING LEFT ONTO HARTFORD AVE. V2 OPERATOR STATED SHE SAW A POLICE OFFICER CRUISER WITH LIGHTS AND SIRENS ON AND STOPPED HER VEHICLE TO ALLOW THE OFFICER TO PASS. V1 OPERATOR STATED THAT HE DID NOT NOTICE THAT V2 STOPPED UNTIL IT WAS TOO LATE, AND REAR ENDED V2. OFFICER [REDACTED] STATED THAT HE APPROACHED THE INTERSECTION WITH EMERGENCY LIGHTS AND SIRENS ACTIVATED AND CAME TO A COMPLETE STOP AT THE RED LIGHT. WHILE STOPPED, OFC. [REDACTED] STATED THAT HE OBSERVED V2 STOP FOR HIM AND THAT V1 HAD ENOUGH TIME AND SPACE TO STOP BEFORE COLLIDING WITH V2. V1 DID NOT STOP AND COLLIDED WITH V2. OFC. [REDACTED] STATED OPERATOR OF V1 WAS AT FAULT. V1 HAS MODERATE DAMAGE TO THE FRONT OF THEIR VEHICLE. V2 HAS MINOR DAMAGE TO THE REAR OF THEIR VEHICLE. NO INJURIES WERE REPORTED. BOTH VEHICLES WERE ABLE TO DRIVE AWAY FROM THE SCENE.

Police car siren & lights → V2 stops vehicle → V1 fails to notice stop → V1 does not brake in time → Rear-end collision → Vehicle damages (no injuries)

LLMs/VLMS for Crash Reports Interpretation



Two-Stage Framework



Dataset



Source

2,033 traffic crash reports from MassDOT in 2022



Focus Area

Accidents occurred at or near **signal-controlled ramp terminals**.



Coverage

Reports span **16 cities**, including *Andover, Bedford, Belmont, Boston*, and others.

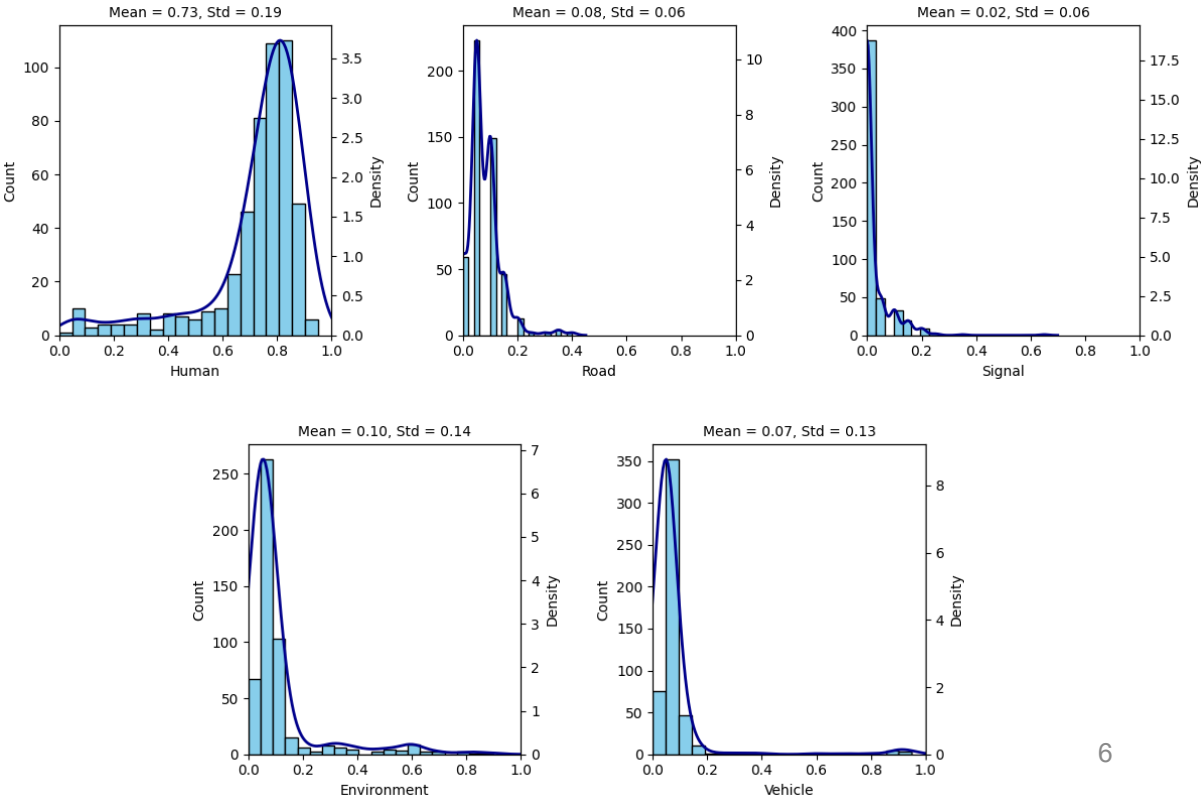


Data Contents

Time and location of each crash; **Crash narrative** (descriptive text explaining the event); **Crash diagram** (visual representation of the accident); **Additional structured details** (e.g., weather, road conditions)

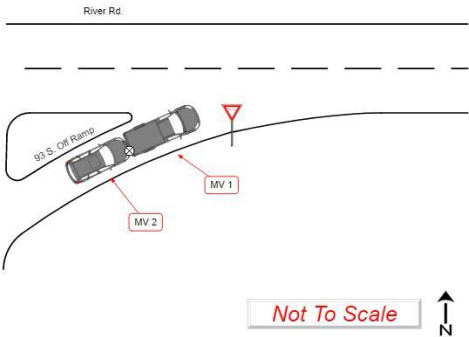
Manually cross-labeled the breakdown of accident root cause proportions in 500 accident reports.

Human	Road	Traffic Signal	Environment	Vehicle
73%	8%	2%	10%	7%



Stage 1: Cause Breakdown

Crash Diagrams



Crash Narratives

On 11/21/22 at approximately 07:37 hrs, MV 2 was exiting 93 South when it rear-ended MV 1, which was also exiting 93 South but had stopped, waiting to yield onto River Rd./Elm St. Auto was detailed for the tow of MV 2, and there were no injuries. Damage is estimated to be greater than \$1,000 to both vehicles.

Factor Decomposition Prompt

You are a traffic safety expert specializing in accident analysis. Your task is to analyze the **given accident report text and accident scene image (if provided)** to determine the root causes of the accident. You must quantify the influence of five key contributing factors.

Contributing Factors

1. Human Factors (HF) – Driver-related errors such as distraction, misjudgment, speeding, fatigue, or failure to yield.
2. Road Factors (RF) – Issues related to lane width, intersection design, sight distance, pavement conditions, or the absence of necessary road infrastructure.
3. Traffic Signal Factors (TSF) – Problems with traffic lights, unclear signals, missing left-turn signals, or improper signal timing.
4. Environmental Factors (EF) – Conditions like rain, snow, fog, glare, or poor lighting that may have contributed to the accident.
5. Vehicle Factors (VF) – Mechanical failures such as brake failure, tire issues, or large vehicle blind spots.

Instructions

- Only return the following structured output in plain text. No additional explanation is needed.
- The percentages must add up to 100%.
- Use the exact format below:
Human Factors: XX%
Road Factors: XX%
Traffic Signal Factors: XX%
Environmental Factors: XX%
Vehicle Factors: XX%



Proportion of Root Cause

Human

Road

Environment

Vehicle

Traffic Signal

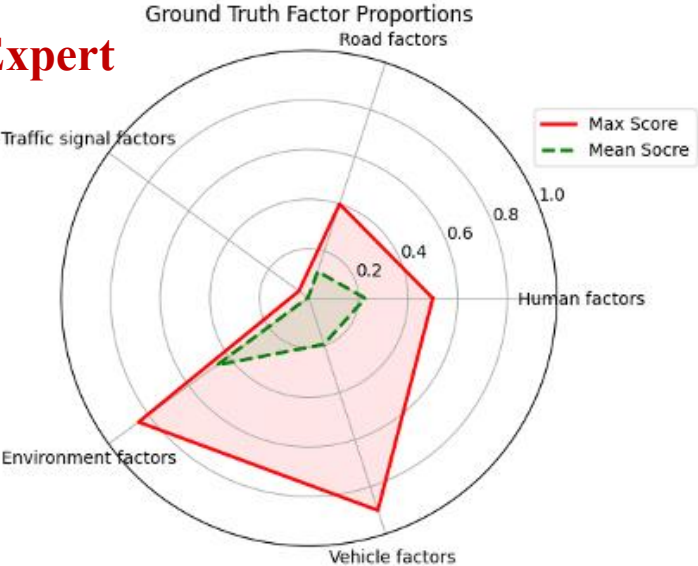
Stage 1: Results Comparison, Which Model Looks More Like an Expert

Local

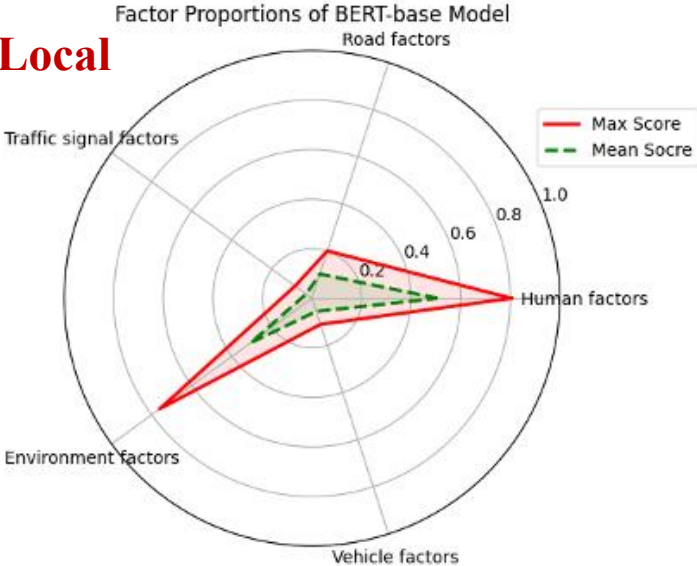
LLM

Model	Input	Spearman	Top-1	Top-3	JSD	WSP
ResNet18+BERT-based	Diagram, Narrative	0.736	0.920	0.747	0.185	0.924
BLIP-2	Diagram, Narrative	0.746	0.920	0.750	0.184	0.923
BERT-based	Narrative	0.737	0.960	0.740	0.170	0.943
GPT-4o	Diagram, Narrative	0.797	0.980	0.813	0.182	0.952
GPT-4o	Narrative	0.799	0.980	0.790	0.178	0.953
Grok 2	Diagram, Narrative	0.863	1.000	0.803	0.131	0.968
Grok 2	Narrative	0.828	1.000	0.793	0.158	0.961
Gemini 2.0 Flash	Narrative	0.736	0.970	0.790	0.193	0.943
DeepSeek V3	Narrative	0.766	0.970	0.750	0.175	0.951

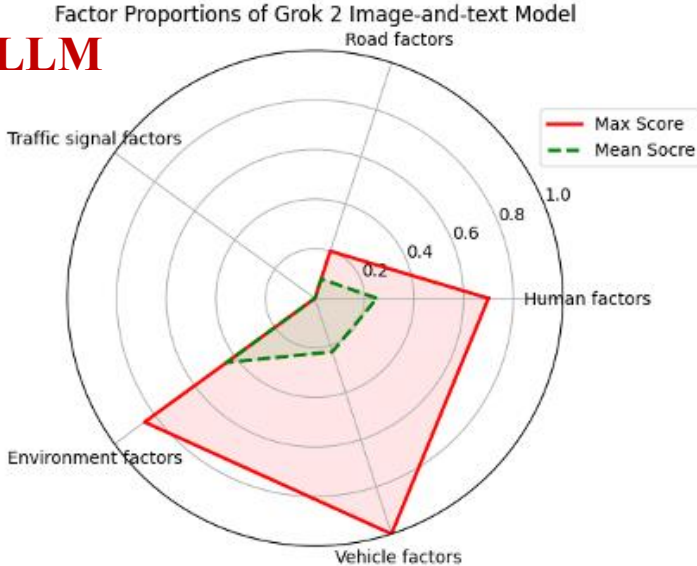
Expert



Local

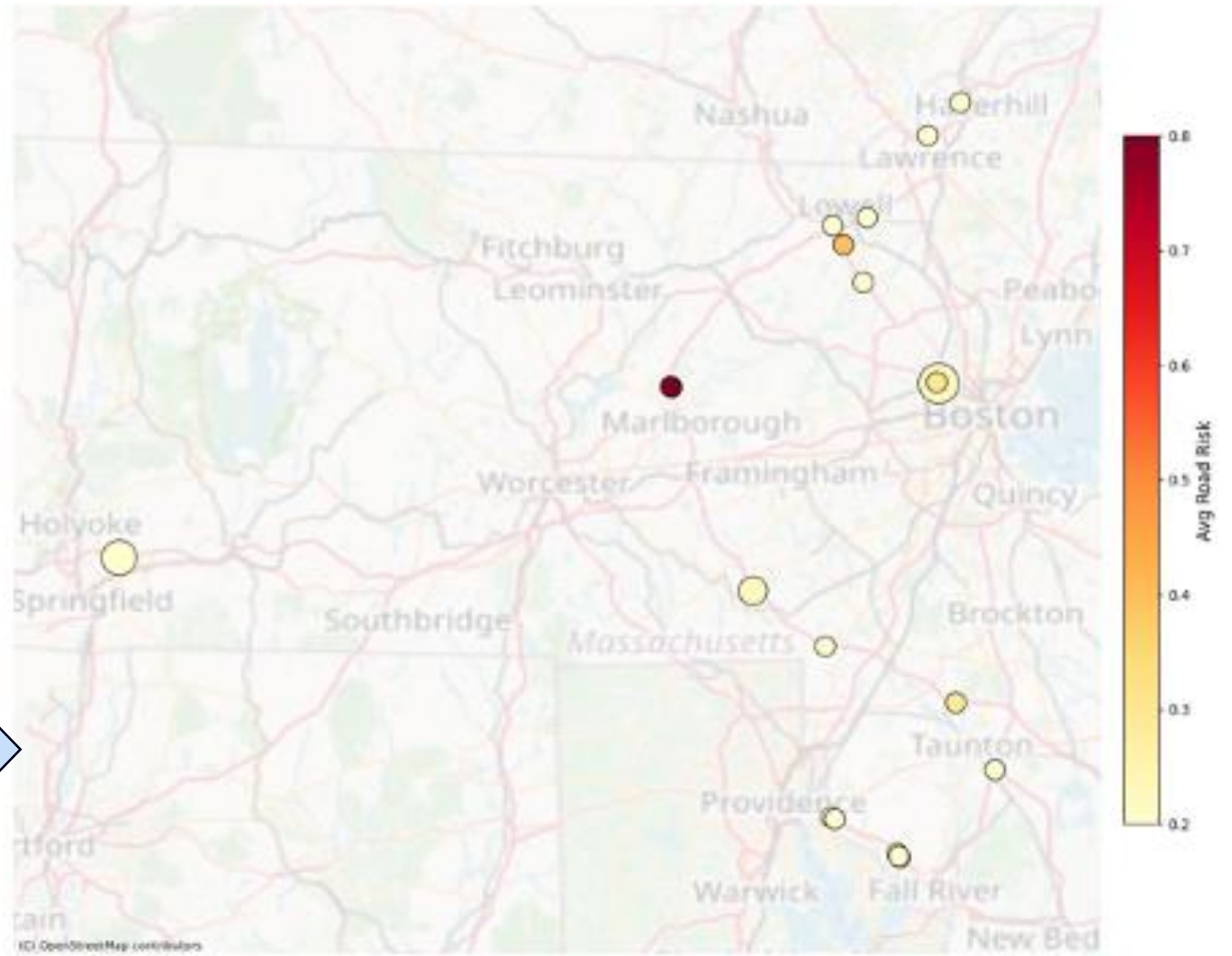


LLM



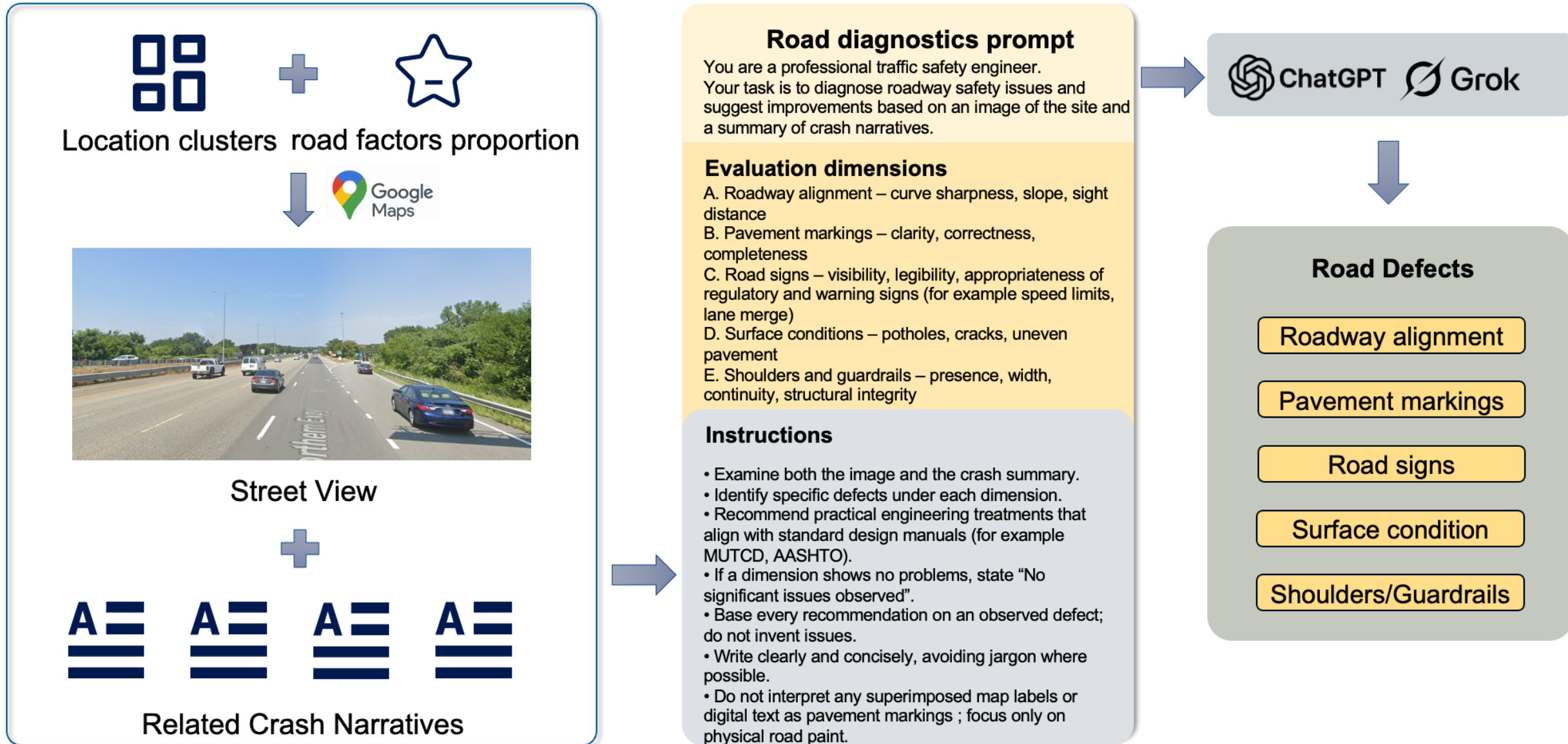
Stage 2: Clustering for High-Risk Road Locations

Geographic Distribution of All 2,033 Crashes Top 20 High-Risk Locations Based on Average Road Attribution Scores



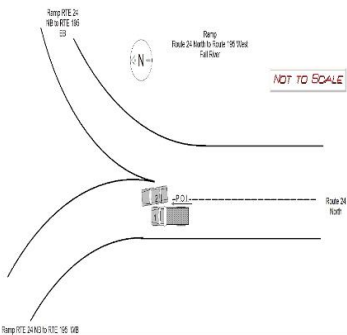
Spatial Clustering Based on Geographic Proximity

Stage 2: Street View Diagnosis – AI as a Road Safety Auditor



Stage 2: Case Study – From Crash to Diagnosis

Input



Crash Narratives

Vehicle 1 is in the left lane and Vehicle 2 is in the right lane. Both vehicles are on Route 24 North and are taking the ramp to Route 195 West. Both left and right lanes are able to take the ramp, and as the ramp begins, a single lane forms. It should be noted there is no signage regarding a single lane forming or which lane ends. As Vehicle 1 and Vehicle 2 enter the single lane portion of the ramp, the two vehicles sideswipe. This causes damage to the driver side of Vehicle 2 and no damage to Vehicle 1. No injuries are reported and both vehicles are able to be driven from the scene. No citation is issued as it could not be determined which vehicle's actions caused the crash due to lack of signage.

Output

Road Diagnostics by Grok 2

Roadway Alignment

Observation: The image shows a curve where the lanes merge into a single lane for the ramp. There are no visible issues with curve sharpness, slope, or sight distance.

Pavement Markings

Observation: The pavement markings do not clearly indicate which lane ends or how the lanes merge into a single lane.

Recommendation: Add clear lane reduction markings that show the right lane tapering off and merging into the left lane. Use dashed lines for the merging lane and solid lines to indicate the continuation of the lane, as suggested by the MUTCD.

Road Signs

Observation: The crash summary indicates there is no signage to inform drivers that the two lanes will merge into one. The image shows a sign for Route 195 West, but there's no sign regarding the lane merge or lane ending.

Recommendation: Install a "Lane Ends Merge Left" sign (MUTCD W4-2) before the merge point on the right lane, and possibly a "Merge" sign (MUTCD W4-1) further back to alert drivers earlier. This will comply with MUTCD standards for warning drivers of lane reductions and merges.

Surface Conditions

Observation: The pavement appears to be in good condition with no visible potholes, cracks, or uneven surfaces.

Recommendation: No significant issues observed.

Shoulders and Guardrails

Observation: The image shows a concrete barrier on the left and a guardrail on the right. Both seem to be in place and structurally sound.

Recommendation: No significant issues observed.

Summary of Recommendations

1. Add lane reduction pavement markings to indicate the merging of lanes.
2. Install appropriate warning signs for lane reduction and merging to ensure drivers are aware of the upcoming change in lane configuration.

Real-world Modification

Before Crash



After Crash

Conclusion, Limitations & Future Work

01

Conclusion

- A new framework powered by large models
- Extracts causal chains from crash reports
- Quantifies root cause proportions
- Clusters crashes to locate high-risk spots
- Diagnoses with street view imagery and provides engineering-based recommendations
- **Results are fast, accurate, and consistent with expert judgment**

02

Limitations

- Street view imagery is **static and constrained**
- Cannot capture **time-sensitive causes** (e.g., temporary obstacles, debris, lighting)
- Fixed viewpoints lead to **incomplete coverage** (often one direction only)
- Diagnostic accuracy requires **multi-perspective street view, LiDAR point clouds, and UAV imagery**

03

Future Work

- Use **LiDAR point clouds** and **drone imagery** to complement street view
- Provide **more complete and accurate diagnostics**
- Extend to **large-scale roadway safety audits**
- Automatically detect safety features: sidewalks, crosswalks, lane configs, roadside obstructions
- **From crash analysis to proactive safety planning**



Thank you

Guanhe Wu

guanhe_wu@student.uml.edu

Xintong Yan

Xintong_yan@uml.edu

Yuanchang Xie

Yuanchang_xie@uml.edu