Collision Patterns and Reporting Blind Spots in 971 California Autonomous Vehicle Crash Reports

MD SHADAB ALAM, Eindhoven University of Technology, The Netherlands
LINGHAN ZHANG, Eindhoven University of Technology, The Netherlands
JIAHUI LI, University of Georgia, United States
FEI DOU, University of Georgia, United States
PAVLO BAZILINSKYY, Eindhoven University of Technology, The Netherlands

Autonomous vehicle (AV) collision reports provide a rare public record of how autonomous driving systems behave in mixed traffic, but their value is limited by heterogeneous layouts, checkbox fields and free-text narratives. This work analyses 971 California Department of Motor Vehicles AV collision reports from October 2014 to March 2026 using a large language model based extraction pipeline and rule based post processing. The analysis identifies recurring scenario classes and evaluates what the reports make visible or leave underspecified. Three patterns dominate the corpus: rear-end collisions involving a stopped AV (267, 27.50%), intersection lateral conflicts (180, 18.54%) and lane-change or merge conflicts (156, 16.07%). The reports support broad scenario classification better than detailed interaction reconstruction: the mean coarse context score was 0.97, whereas the mean fine context score was 0.48. The largest blind spots concerned speed, lane position and traffic context. These findings suggest that many reported AV crashes occur in ordinary but interaction-heavy traffic situations where expectation, timing and coordination are difficult to align.

Additional Key Words and Phrases: Autonomous Vehicles, Collision Reports, Mixed Traffic, Crash Taxonomy, Large Language Models

1 Introduction

Autonomous vehicles (AVs) are expected to improve road safety and transport efficiency, yet these benefits depend on safe operation in mixed traffic where automated systems interact with manually driven vehicles, cyclists, pedestrians and other road users [15, 17, 18]. These interactions are not only technical perception and control problems. They also involve expectations, informal negotiation and communication cues that may be weakened when a human driver is absent or visually disengaged [6, 12]. AV safety therefore needs to be studied through the everyday interaction situations in which autonomous and human road users meet.

Public collision reports are one of the few sources that document real-world AV incidents outside controlled experiments. California has required manufacturers to report qualifying AV collisions since the start of its testing programme, and the resulting reports combine structured fields, marked checkboxes, vehicle damage diagrams and narrative descriptions [3, 4]. Previous studies have shown that many reported AV collisions are low-severity events, often involving rear-end impacts and non-AV parties [5, 10, 11]. Topic modelling and statistical analyses have further identified themes such as stopped AVs, manual transitions, overtaking conflicts and spatial clustering in dense urban environments [2, 14].

However, the reporting format also constrains what can be learned. Many earlier analyses rely on manually coded variables or predefined categories, while the reports themselves contain information distributed across layout, check-boxes and prose. Recent work has shown that large language models can support crash-report interpretation and other road-user reasoning tasks [1, 9, 13]. This makes it possible to treat the reports not simply as a set of existing form fields,

Authors' Contact Information: Md Shadab Alam, m.s.alam@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands; Linghan Zhang, l.zhang1@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands; Jiahui Li, jiahui.li1@uga.edu, University of Georgia, Athens, United States; Fei Dou, fei.dou@uga.edu, University of Georgia, Athens, United States; Pavlo Bazilinskyy, p.bazilinskyy@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands.

but as visually structured documents from which scenario-level patterns and reporting blind spots can be extracted at scale.

1.1 Aim of Study

The aim of this study is to examine traffic collisions involving AVs in California by combining large scale report processing with an interaction centred interpretation of the resulting data. Although recent research has analysed autonomous vehicle crashes, these studies focus primarily on visualising descriptive statistics [16]. Rather than treating collision reports only as a source of isolated variables, we use a language model based coding pipeline to derive a structured representation of publicly available reports, identify recurring incident scenarios, and examine what the reports reveal and fail to reveal about collisions involving AVs in mixed traffic. In doing so, the study contributes an empirical taxonomy of recurring collision scenarios, an assessment of reporting blind spots, and a methodological basis for analysing AV collision reports at scale.

2 Method

2.1 Data Source and Corpus Construction

On 4 April 2026, we collected 971 publicly available AV collision reports from the California Department of Motor Vehicles (DMV) portal (https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-collision-reports; on 28 April 2026, DMV removed these reports from public access and UC Berkely introduced Autonomous Vehicle Safety dashboard https://tims.berkeley.edu/tools/avsafety.php). The reports with recoverable dates span from 14 October 2014 to 18 March 2026. Manufacturer information was recoverable for 969 reports. After normalising legacy and related names, the corpus was dominated by Waymo (450 cases), Cruise (241), Zoox (167), Apple Inc. (23) and Pony.ai (15). The remaining cases were distributed across smaller reporting entities. Each report documents a collision involving a permitted AV, but these are manufacturer-submitted regulatory documents rather than police reports written at the crash scene. The OL 316 form records crash date and location, vehicle details, damage, injuries, movement preceding collision, collision type, environmental conditions and a narrative description. Figure 1 shows the report structure used by the extraction pipeline.

2.2 Prompting Procedure and Response Collection

We used GPT-5.4 Thinking[1] as a document analysis tool under a fixed prompting protocol. For each report, the model received the collision report PDF and a query document with 25 questions (Q0–Q24). The questions covered responsibility, manufacturer and vehicle information, crash date and location, road context, damage, injury, other-party information, AV mode, movement, lane position, intersection involvement, direction of travel, weather, lighting, roadway surface, roadway condition, movement preceding collision, type of collision and associated factors.

The prompt instructed the model to treat the AV as Vehicle 1 and the other involved road user as Vehicle 2 or Other Party, to answer only from the relevant report section, and to return `NA` when confidence was below 50%. The prompts were refined through pilot tests before processing the full corpus. Model responses were recorded verbatim and stored in tabular form for parsing.

[1] One report, `Waymo-OL316-032224-Amended-Redacted.pdf`, was analysed using GPT-5.4 with high reasoning effort because GPT-5.4 Thinking was not available. OpenAI describes GPT-5.4 as a frontier model for complex professional work, and its API documentation lists `high` as a supported reasoning-effort setting, so we treat this run as the closest available equivalent to GPT-5.4 Thinking.

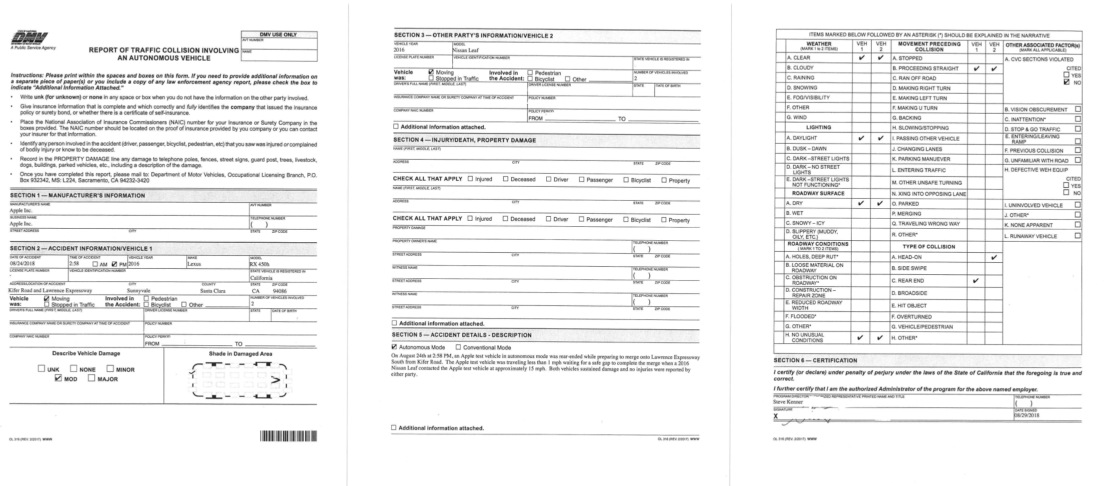

Fig. 1. Example autonomous vehicle collision report from the California Department of Motor Vehicles corpus used in this study (Apple Inc., 24 August 2018; file: `Apple_082418.pdf`). The three page PDF illustrates the document structure processed by our pipeline: page 1 contains manufacturer and crash metadata together with a vehicle damage diagram, page 2 contains the other party information and narrative crash description and page 3 contains checkbox based fields for weather, lighting, roadway surface, roadway conditions, movement preceding collision, type of collision and other associated factors.

2.3 Post Processing and Derived Analytical Variables

The recorded responses were converted to a canonical schema using rule based post processing. The pipeline normalised field names and values, retained raw responses for traceability, and derived grouped variables for road-user type, AV movement, other-party movement, collision group, blame group, intersection context, environmental profile and scenario class. Scenario classes were assigned using explicit combinations of movement, collision group, intersection context, road-user vulnerability and cues from the explanation text.

The taxonomy included *AV stopped rear end*, *intersection lateral conflict*, *lane change or merge conflict*, *turn across path conflict*, *curbside or parked vehicle conflict*, *low speed stop or obstruction case*, *vulnerable road user interaction* and *other or ambiguous*. To assess report quality, fields were grouped by provenance: bounded form fields, checkbox fields, narrative fields and online enriched fields. We then calculated report completeness, explicitness, coarse context, fine context and a context granularity gap. Missing values were retained as `NA`, rather than removed, because missingness itself was central to the blind-spot analysis [7].

2.4 Analytical Strategy

The analysis was organised in two stages. First, we used the structured dataset to identify recurring collision scenarios and their relation to road-user type, movement, environmental profile and reported attribution. Secondly, we analysed the reporting process itself by examining missingness, provenance, scenario determinability, movement consistency and the difference between coarse and fine context. Figure 2 gives a corpus-level overview of the main analytical dimensions.

3 Results

The final corpus contained 971 report entries. All had a usable model response and were parsed successfully. Across the corpus, the mean report explicitness score was 0.79, indicating that much of the form-level information could be recovered, although not all field types were equally supported. The mean coarse context score was 0.97, while the mean fine context score was 0.48, giving a mean context granularity gap of 0.50. In other words, the reports captured broad scenario structure much more consistently than detailed interaction context.

The scenario distribution was concentrated in a small number of recurring patterns. The largest scenario class was *AV stopped rear end* with 267 cases (27.50%), followed by *other or ambiguous* with 212 cases (21.83%), *intersection lateral conflict* with 180 cases (18.54%) and *lane change or merge conflict* with 156 cases (16.07%). The remaining classes were *curbside or parked vehicle conflict* with 116 cases (11.95%), *vulnerable road user interaction* with 17 cases (1.75%), *turn across path conflict* with 14 cases (1.44%) and *low speed stop or obstruction case* with 9 cases (0.93%). This pattern is shown in Figure 3.

The other involved party was classified as a vehicle in 902 cases (92.89%), unknown in 46 cases (4.74%), a cyclist in 14 cases (1.44%) and a pedestrian in 3 cases (0.31%). Reported blame was assigned to the other road user in 805 cases (82.90%), the AV as primary party in 136 cases (14.01%), environmental or road conditions in 28 cases (2.88%) and unclear in 2 cases (0.21%). Figure 4 shows that these accountability patterns varied across scenario classes rather than being distributed uniformly.

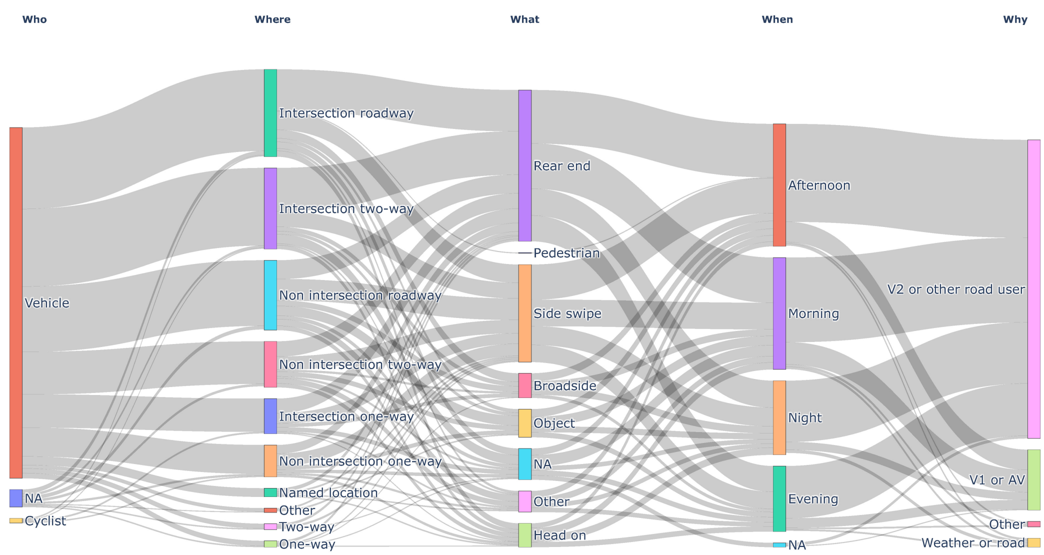

Fig. 2. Sankey diagram showing how collision cases are distributed across the main analytical dimensions in the dataset.

The reports were substantially more informative for coarse scenario structure than for fine-grained interaction detail. Missingness was highest for *v2 speed* (87.95%), *v2 lane* (78.06%), *street busy* (77.75%), *v1 lane* (76.52%) and *v1* *speed* (75.59%). Missingness was also substantial for *v2 damage description* (52.52%), *lane number* (49.74%) and *street type* (44.08%). Field availability differed across provenance groups, with the highest availability for bounded form fields (0.85), followed by checkbox fields (0.76), narrative fields (0.70) and online enriched fields (0.49). Scenario determinability was high in 721 cases (74.25%), medium in 241 cases (24.82%) and low in 9 cases (0.93%). Movement consistency was much weaker: 687 reports (70.75%) were flagged as inconsistent or ambiguous at the movement-detail level. As shown in Figure 5, the most severe blind spots were concentrated in the variables needed for interaction reconstruction.

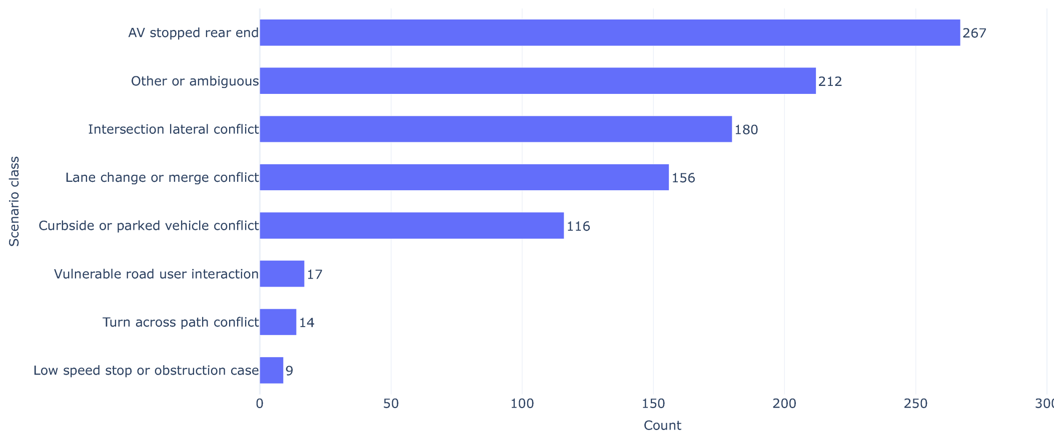

Fig. 3. The distribution of scenario classes in the collision corpus. Each bar represents the number of cases assigned to a given scenario class, allowing comparison of the most common recurring collision situations, including AV stopped rear end crashes, intersection lateral conflicts, lane change or merge conflicts, curbside or parked vehicle conflicts, turn across path conflicts, vulnerable road user interactions, low speed stop or obstruction cases and cases that remained other or ambiguous.

4 Discussion

The results indicate that collisions involving AVs in California are concentrated in a limited set of recurring situations rather than being evenly spread over many unrelated crash types. The most prominent patterns were *AV stopped rear end*, *intersection lateral conflict*, *lane change or merge conflict* and *curbside or parked vehicle conflict*. As shown in Figures 2 and 3, these scenario classes account for most of the corpus, although a substantial minority of reports remained *other or ambiguous*. Taken together, this pattern suggests that many reported AV collisions arise in common mixed-traffic situations in which prediction, coordination and timing are especially important.

The prominence of stopped AV rear-end cases is especially important. In many such cases, the AV may have behaved conservatively or legally, yet the following driver may not have anticipated the timing or extent of the stop. This points to a boundary between formal rule compliance and social predictability in traffic. Similarly, intersection lateral conflicts and lane-change or merge conflicts are coordination problems in which small differences in expectation can become safety relevant. The results therefore support an interaction-centred view of AV safety, where crashes are studied as breakdowns in shared interpretation and coordination rather than as isolated technical failures.

A second contribution is the distinction between broad scenario knowledge and fine interaction detail. The high coarse context score shows that the reports can identify what generally happened. The lower fine context score shows that they often cannot explain exactly how the interaction unfolded. Lane position, speed, road busyness and

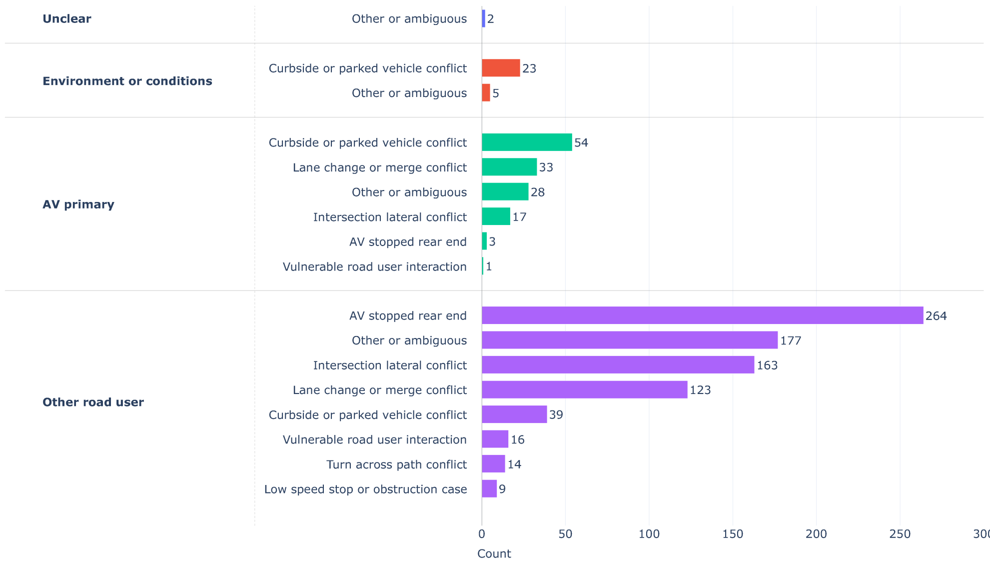

Fig. 4. The distribution of blame attribution across scenario classes in the collision corpus. For each scenario class, the bars indicate how many cases were assigned to the AV, the other road user, environmental or road condition factors, or remained unclear, allowing comparison of accountability patterns across different types of collision situations.

detailed movement information were often missing or ambiguous. This matters because those variables are central to understanding whether a crash resulted from cautious AV behaviour, unexpected human behaviour, unusual traffic geometry, insufficient communication, or a combination of these factors.

The provenance results show that the reporting form shapes the analysis. Structured form fields and checkboxes are easier to recover than information that depends on prose or external lookup. Narrative descriptions remain valuable, but they vary in detail and framing. Online enrichment can add context, but it introduces uncertainty because it is not part of the original report. These findings suggest that current reports are useful for detecting recurring patterns and blind spots, but should be used carefully when making claims about detailed causality.

For automotive user interface and interaction research, these blind spots are important because the missing fields correspond to the variables that would usually explain coordination. A report may state that an AV was stopped, turning, changing lanes or proceeding through an intersection, but still omit the relative speed, lane position, traffic density or sequence of actions that made the interaction intelligible to the other road user. The taxonomy should therefore be seen as a way of organising reported interaction patterns and identifying reporting gaps, rather than as a substitute for naturalistic safety analysis or detailed crash reconstruction.

4.1 Limitations and future work

This study is limited to California DMV collision reports and therefore reflects one regulatory context, one reporting infrastructure and the operating conditions of permitted AV testing in that jurisdiction. The reports are manufacturer-submitted documents rather than independent crash reconstructions, and should not be treated as ground truth records of the event. They may omit relevant details, simplify the sequence of actions or frame responsibility in ways shaped by reporting practice. The corpus also does not contain an exposure denominator, such as distance driven, operational

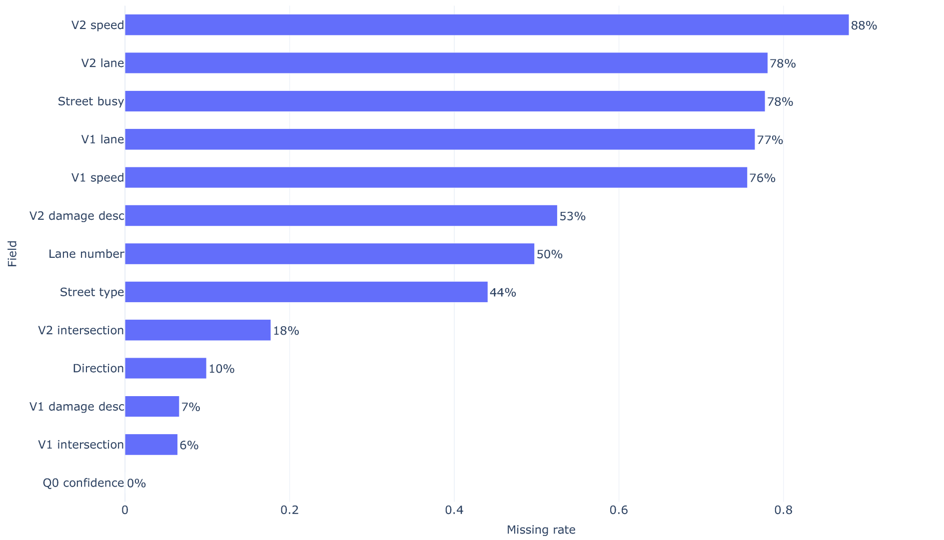

Fig. 5. Field-level missingness for selected blind-spot variables in the collision report corpus. The figure shows the proportion of reports in which fine-grained contextual variables, including speed, lane position, street busyness, lane number, street type, intersection detail, direction and damage descriptions, were unavailable after extraction. Higher missingness rates indicate dimensions of crash reconstruction that are systematically under-specified in the source material. Overall, the pattern shows that the reports support broad scenario interpretation more consistently than detailed interaction-level reconstruction.

design domain, fleet size or traffic volume. The results therefore describe patterns among reported collisions, not comparative crash risk across manufacturers, vehicle platforms or deployment strategies.

A second limitation concerns the information available in the reports. The analysis shows that broad scenario structure was usually recoverable, but the variables most important for interaction reconstruction were often missing. Lane position, relative speed, traffic density, road geometry and the precise order of actions were not consistently documented. This means that some cases could be placed in a scenario class without supporting a detailed causal interpretation. The *other or ambiguous* category should therefore be read as a mixture of unusual cases, incomplete reporting and situations where the available text did not support a more specific assignment.

The language-model based workflow also introduces methodological constraints. Although the prompt was fixed after pilot testing and the post processing was rule based, the extracted fields, scenario classes and blind-spot indicators remain derived variables. The present version does not yet include a full independently hand-coded gold standard or an inter-rater reliability assessment. Online enrichment fields, such as lane count, street type and likely busyness, add contextual value but are less directly grounded in the collision report than form fields or checkbox fields. Future work should therefore validate the extraction on a manually annotated subset, compare multiple document-understanding models, and report agreement separately for form fields, checkbox fields, narrative fields and derived scenario variables.

Future studies should also connect this reporting analysis with richer sources of evidence. Linking DMV reports with disengagement reports, police records, insurance data, map data, simulation logs or naturalistic driving data would help distinguish reporting artefacts from recurring interaction problems. Comparative studies across jurisdictions would clarify whether the same stopped-AV, intersection and lane-change patterns appear under different road rules, reporting formats and cultural expectations. Finally, the findings point to concrete improvements in future reporting forms. A short structured sequence of events, fields for relative motion, lane position, speed, traffic density, intersection layout, AV control state and other road-user intent, and an indication of supporting material such as police reports or video review would make future reports more useful for comparing cases and converting recurring patterns into reusable evaluation scenarios.

A direction is to turn the scenario taxonomy into a bridge between retrospective crash analysis and prospective evaluation. The dominant classes identified here could be used to define reusable test situations for simulation, closed-track studies and human-in-the-loop experiments. For example, stopped-AV rear-end cases could motivate studies of how following drivers interpret cautious autonomous behaviour, while intersection and lane-change conflicts could be translated into scenarios that vary priority, timing, gap size and external communication. Such work would help separate what is a limitation of the report from what is a recurrent interaction challenge. It would also make the findings more actionable for interface design by linking reported crashes to specific questions about how AVs communicate intent, how other road users infer that intent and which contextual cues should be captured in future reporting.

5 Supplementary Material

In line with current open science practices and recommendations for transparency in automotive research [8], the authors openly provide these research artefacts to support reproducibility, collaboration and further advancements in the field. The analysis code, query and responses of LLMs are available at https://www.dropbox.com/scl/fo/k7kqrj86x0ib5n4x6ttab/ADwuUh3Xex8TjfEIdt3sdEo?rlkey=48l0n7zxzlni10qxpb6ogv2dj&st=a7tr6uav. A maintained version of the code is available at https://github.com/bazilinskyy/llm-events.

References

[1] Md Shadab Alam and Pavlo Bazilinskyy. 2025. Cross or Nah? LLMs Get in the Mindset of a Pedestrian in front of Automated Car with an eHMI. In *Adjunct Proceedings of the 17th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 19–34. doi:10.1145/3744335.3758477

[2] Hananeh Alambeigi, Anthony D McDonald, and Srinivas R Tankasala. 2020. Crash themes in automated vehicles: A topic modeling analysis of the California Department of Motor Vehicles automated vehicle crash database. *arXiv preprint arXiv:2001.11087* (2020).

[3] California Department of Motor Vehicles. 2026. Autonomous Vehicle Collision Reports. Accessed 2026-04-07. https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-collision-reports/

[4] California Department of Motor Vehicles. 2026. Autonomous Vehicles Testing with a Driver. Accessed 2026-04-07. https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/testing-autonomous-vehicles-with-a-driver/

[5] Subasish Das, Anandi Dutta, and Ioannis Tsapakis. 2020. Automated vehicle collisions in California: Applying Bayesian latent class model. *IATSS research* 44, 4 (2020), 300–308. doi:10.1016/j.iatssr.2020.03.001

[6] Debargha Dey, Azra Habibovic, Andreas Löcken, Philipp Wintersberger, Bastian Pfleging, Andreas Riener, Marieke Martens, and Jacques Terken. 2020. Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. *Transportation Research Interdisciplinary Perspectives* 7 (2020), 100174. doi:10.1016/j.trip.2020.100174

[7] Yiran Dong and Chao-Ying Joanne Peng. 2013. Principled missing data methods for researchers. *Springerplus* 2, 1 (2013), 222.

[8] Patrick Ebel, Pavlo Bazilinskyy, Mark Colley, Courtney Michael Goodridge, Philipp Hock, Christian P. Janssen, Hauke Sandhaus, Aravinda Ramakrishnan Srinivasan, and Philipp Wintersberger. 2024. Changing lanes toward open Science: Openness and transparency in automotive user research. In *Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Stanford, CA, USA) *(AutomotiveUI '24)*. Association for Computing Machinery, New York, NY, USA, 94–105. doi:10.1145/3640792.3675730

[9] Mohammad El-Yabroudi, Sri Harsha Pothuguntla, Athar Ghadi, and Balakumar Muniandi. 2025. Harnessing Generative AI for Text Analysis of California Autonomous Vehicle Crashes OL316 (2014–2024). *Electronics* 14, 4 (2025), 651. doi:10.3390/electronics14040651

[10] Francesca M Favarò, Nazanin Nader, Sky O Eurich, Michelle Tripp, and Naresh Varadaraju. 2017. Examining accident reports involving autonomous vehicles in California. *PLoS one* 12, 9 (2017), e0184952. doi:10.1371/journal.pone.0184952

[11] Noah J Goodall. 2021. Comparison of automated vehicle struck-from-behind crash rates with national rates using naturalistic data. *Accident Analysis & Prevention* 154 (2021), 106056. doi:10.1016/j.aap.2021.106056

[12] Daniël D Heikoop, Marjan Hagenzieker, Giulio Mecacci, Simeon Calvert, Filippo Santoni De Sio, and Bart Van Arem. 2019. Human behaviour with automated driving systems: a quantitative framework for meaningful human control. *Theoretical issues in ergonomics science* 20, 6 (2019), 711–730. doi:10.1080/1463922X.2019.1574931

[13] Maroa Mumtarin, Md Samiullah Chowdhury, and Jonathan Wood. 2023. Large Language Models in Analyzing Crash Narratives–A Comparative Study of ChatGPT, BARD and GPT-4. *arXiv preprint arXiv:2308.13563* (2023).

[14] Ronik Ketankumar Patel, Sai Sneha Channamallu, Muhammad Arif Khan, Sharareh Kermanshachi, and Apurva Pamidimukkala. 2024. An exploratory analysis of temporal and spatial patterns of autonomous vehicle collisions. *Public Works Management & Policy* 29, 4 (2024), 588–611. doi:10.1177/1087724X231217677

[15] Amir Rasouli and John K Tsotsos. 2019. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE transactions on intelligent transportation systems* 21, 3 (2019), 900–918. doi:10.1109/TITS.2019.2901817

[16] Transport Research Centre (CDV). 2026. Statistics. AV Crashes. Accessed: April 7, 2026. https://www.avcrashes.net/stats

[17] Jun Wang, Li Zhang, Yanjun Huang, and Jian Zhao. 2020. Safety of autonomous vehicles. *Journal of advanced transportation* 2020, 1 (2020), 8867757. doi:10.1155/2020/8867757

[18] Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. 2020. A survey of autonomous driving: Common practices and emerging technologies. *IEEE access* 8 (2020), 58443–58469. doi:10.1109/ACCESS.2020.2983149

Appendix

A Extraction Protocol

This appendix summarises the additional methodological material made available with the supplementary artefacts. During extraction, the AV was treated as Vehicle 1 and the other involved road user as Vehicle 2 or Other Party. The prompt restricted answers to evidence in the report, requested compact responses, and required `NA` when the report did not support a confident answer. This was necessary because the source documents combine form fields, checkboxes, damage diagrams and narrative descriptions rather than a single plain-text record.

The query document contained 25 questions, Q0–Q24. Q0 asked for a responsibility judgement, explanation, main contributing factor and confidence estimate. Q1–Q3 extracted manufacturer, vehicle, date, time and location fields. Q4–Q6 added road context from address-based lookup. Q7–Q17 covered damage, injury, other-party type, AV operating mode, movement, lane position, intersection involvement, direction of travel and damage descriptions. Q18–Q24 focused on checkbox-based fields, including weather, lighting, roadway surface, roadway condition, movement preceding collision, collision type and associated factors. The full query document, extracted responses and analysis outputs are provided through the supplementary material.

B Derived Variables and Diagnostics

The extracted responses were converted into a structured dataset and grouped into analysis-facing variables: road-user type, vulnerability, AV mode, AV and other-party movement, collision group, blame group, intersection context, environmental friction profile, harm scope and scenario class. Scenario classes were assigned from explicit combinations of movement, collision group, intersection context, road-user vulnerability and explanatory cues. For each case, the diagnostic output retained the scenario trigger, evidence source and evidence count so that broad assignments could be distinguished from weaker cases.

The diagnostics also describe what the reports make visible and what they leave underspecified. Fields were grouped by provenance as form, checkbox, narrative and online-enriched fields, and these groups were used to calculate availability, explicitness, coarse context, fine context and the context granularity gap. Additional checks recorded determinability, movement consistency and blame-confidence alignment.