Collision Patterns and Reporting Blind Spots in 970 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 collision reports offer a rare view of how autonomous driving systems perform in mixed traffic, but they are difficult to analyse at scale because they combine structured form fields, checkbox-marked elements, and free-text narratives. We analysed 970 publicly available California Department of Motor Vehicles collision reports dated between October 2014 and March 2026, using a ChatGPT 5.4 Thinking extraction pipeline to derive a structured dataset for scenario-level analysis. Three recurring patterns dominated the corpus: rear-end crashes involving stopped AV (266, 27.4%), intersection lateral conflicts (180, 18.6%) and lane change or merge conflicts (156, 16.1%). The reports supported coarse scenario structure much better than fine-grained interaction reconstruction, with mean coarse and fine context scores of 0.97 and 0.48, respectively. Taken together, the findings suggest that many reported AV collisions occur in mixed traffic situations where prediction, coordination, and road user expectations are difficult to align.

Additional Key Words and Phrases: Autonomous vehicles, Traffic collision reports, Collision taxonomy, Mixed traffic, Safety analysis

ACM Reference Format:
Md Shadab Alam, Linghan Zhang, Jiahui Li, Fei Dou, and Pavlo Bazilinskyy. 2026. Collision Patterns and Reporting Blind Spots in 970 California Autonomous Vehicle Crash Reports. 1, 1 (April 2026), 16 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Autonomous vehicles (AVs) are widely expected to improve road safety, traffic efficiency, travel convenience, and environmental performance of transport systems [58, 61]. However, these expectations depend on the assumption that AVs can operate robustly in the complexity of real world traffic. In practice, AVs do not drive in isolation. They interact with manually driven vehicles (MDVs), cyclists, pedestrians, and other road users whose behaviour is diverse, context dependent, and often difficult to anticipate. Understanding the safety of AV operation for both vehicle occupants and surrounding road users therefore requires attention not only to the technical capabilities of autonomous driving systems, but also to how AVs negotiate intent, priority, and expectations in mixed traffic [29, 48].

Current AVs remain vulnerable to a wide range of technical and operational limitations. Their performance depends on detection systems, perception and prediction algorithms, software reliability, and the ability to deal with unusual or

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, North Brabant, The Netherlands; Jiahui Li, jiahui.li1@uga.edu, University of Georgia, Athens, Georgia, United States; Fei Dou, fei.dou@uga.edu, University of Georgia, Athens, Georgia, United States; Pavlo Bazilinskyy, p.bazilinskyy@tue.nl, Eindhoven University of Technology, Eindhoven, The Netherlands.

Manuscript submitted to ACM

degraded traffic conditions [15, 17]. Well known fatal crashes illustrate how these limitations can arise through different combinations of technical, human and organisational failure [43, 44]. In the 2016 Williston, FL (US) crash, a Tesla operating with Tesla's Autopilot suite, specifically Traffic Aware Cruise Control (TACC) and the Autosteer lane keeping system, corresponding to SAE Level 2 partial driving automation, struck a tractor semitrailer crossing its path, revealing limits in hazard detection along with driver inattention and overreliance on automation [4, 43]. In the 2018 Tempe, AZ (US) crash, a test vehicle controlled by Uber's developmental automated driving system struck and killed a pedestrian, with the subsequent investigation pointing not only to perception and supervision problems, but also to weaknesses in safety risk assessment and operator oversight [39, 44]. Together, these cases show that autonomous driving failures can not arise from a single fault alone, but from the interaction of perception limits, human supervision, and organisational safety practice [39, 43, 44]. The differences in sensing strategies and system design between manufacturers further shape how such limitations appear in practice. Some AV platforms rely on extensive sensor fusion and detailed map support, whereas others prioritise camera based perception and real time learning [33, 54, 57]. These differences may influence how AVs perform in similar situations and how they fail when conditions become challenging.

At the same time, AV safety cannot only be understood through the lens of automation failure. Human road users introduce substantial behavioural variability. Driving behaviour differs between individuals and situations in terms of speed choice, gap acceptance, lane changing, compliance with traffic rules, reaction time, risk taking, and responses to complex traffic events [50, 60]. This heterogeneity is influenced by demographic, psychological, situational, and cultural factors [1, 6, 23, 32, 55]. In mixed traffic, AVs must therefore interpret and respond to behaviour that may be inconsistent, strategic, or ambiguous. The challenge is not simply to perceive the environment, but to negotiate with road users whose intentions are often only partially observable [48].

Interaction in mixed traffic is further shaped by trust, expectation, and communication. Drivers of MDVs can calibrate their behaviour differently when they encounter recognisable AVs [52]. Trust in automation also varies between individuals and can affect attention, reliance, and intervention behaviour [30, 36, 41]. In addition, road users often rely on subtle social cues such as eye contact, gestures, and implicit conventions to communicate intent [28, 40, 62]. These cues are often absent, weakened, or transformed in interactions involving AVs [5, 19]. Consequently, collisions involving AVs may reflect not only sensing or control limitations, but also broader breakdowns in coordination, expectation, and communication in mixed traffic.

For this reason, collision reports involving AVs provide an important empirical resource for understanding how autonomous driving systems perform outside controlled testing environments. In California (US), public collision reporting has been part of the DMV's autonomous vehicle testing framework since 2014, when the state began its Autonomous Vehicle Tester Programme [13]. Under the current DMV framework, manufacturers must report within 10 days any collision involving a vehicle able to drive autonomously that results in property damage, bodily injury, or death [10]. The DMV currently defines an autonomous test vehicle as one whose autonomous technology, when activated, performs the dynamic driving task, while excluding driver assistance systems that require constant human control or active monitoring [12]. These reports combine structured fields, such as vehicle details, damage, and environmental conditions, with free text descriptions of crash circumstances. Previous work has used these reports to identify recurring collision types, severity patterns, and possible contributing factors. In studies, rear end crashes have often been found to be the most frequent type of collision, many reported crashes have involved low speeds, and in many cases the non AV party has been considered fault [7, 18, 24, 25, 38, 59]. Other work has explored themes in crash narratives, such as manual transitions, overtaking conflicts, stopped rear end crashes of AV, and spatial clustering in dense urban environments [3, 14, 46]. Together, these studies demonstrate the value of AV collision reports, but also show that current understanding still depends heavily on manually coded variables, predefined feature sets, or relatively narrow analytical tasks.

Recent advances in generative AI have opened new possibilities for analysing such reports at scale. Large language models (LLMs) have begun to be used to interpret crash narratives, answer structured questions about reports, and compare model outputs across different crash analysis tasks [22, 42]. More broadly, LLMs have also been used to model road user reasoning in autonomous driving contexts, for example, by estimating whether a pedestrian would decide to cross in front of an autonomous car with an external human machine interface [2]. This is relevant because AV safety is not only a matter of sensing and control, but also of interpreting interaction, expectation, and intent. In parallel, vision language models (VLMs) and related document understanding models have shown strong capability in extracting information from visually rich documents that combine text, layout, and graphical structure [31, 35, 37]. This is particularly important for AV collision reports, which contain not only narrative text but also form structure, checkboxes, and layout dependent information that are difficult to process using conventional rule based pipelines alone.

Despite this progress, several limitations remain. First, many existing studies still rely on manually coded or pre selected features, which constrain the analysis to assumptions made in advance. Secondly, the effort required to read and code hundreds of reports makes it difficult to keep up with the growing volume of available data. Third, studies often focus on isolated variables even though collisions emerge from combinations of road user behaviour, environmental context, vehicle state, and interaction dynamics. Existing AI based work has often focused on specific tasks, such as interpreting crash narratives or comparing model answers, rather than on systematically assessing which fields can be extracted robustly, which remain ambiguous, and which aspects of the reports are underspecified. This kind of study is therefore needed because existing work has more often described AV collisions through isolated variables or narrowly defined tasks than through interaction patterns across cases and has paid less attention to what the reporting format itself makes visible or leaves underspecified. The novelty of the present work lies in treating AV collision reports not merely as a source of predefined variables, but as a basis for deriving an empirical taxonomy of recurring scenarios, identifying reporting blind spots and examining the robustness of extracted information at scale [29, 34].

1.1 Aim of the 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 (AV) crashes, these studies focus primarily on visualising descriptive statistics [56]. 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

We collected 970 publicly available autonomous vehicle collision reports from the California Department of Motor Vehicles website on 4 April 2026 and used them as the study corpus. Among the reports with recoverable dates, the

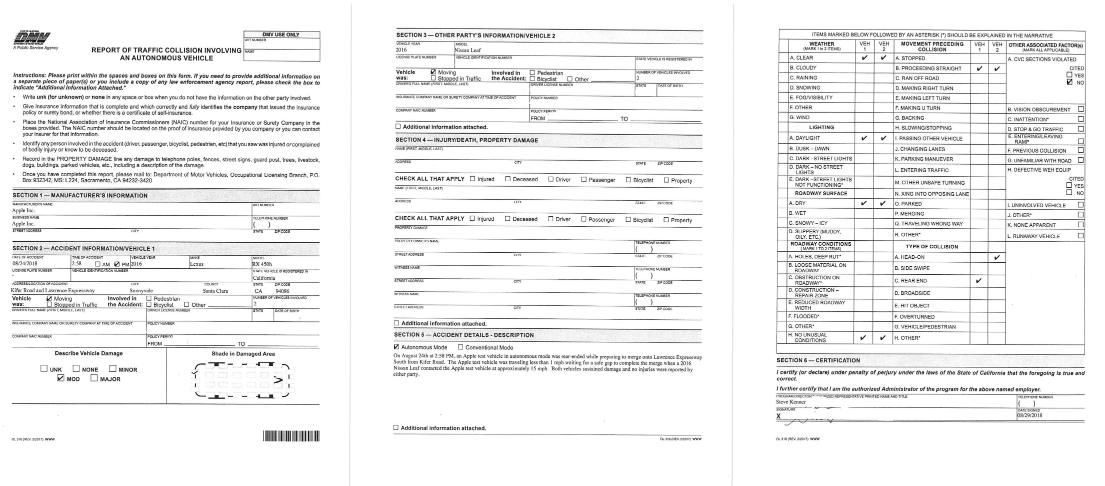

Fig. 1. Example autonomous vehicle collision report from the California Department of Motor Vehicles corpus used in this study (Apple Inc., 24 August 2018; file: `Apple_082418.pdf`). The three page PDF illustrates the document structure processed by our pipeline: page 1 contains manufacturer and crash 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.

corpus spans from 14 October 2014 to 18 March 2026. All reports were downloaded from the *Autonomous Vehicle Collision Reports* portal (https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-collision-reports). AV manufacturer information was recoverable for 968 of the 970 reports. To improve reporting consistency, manufacturer names were normalised and, where appropriate, consolidated under the latest name when different labels referred to the same reporting entity or company lineage.[1] After normalisation, the five most frequent manufacturer labels were Waymo (449 cases), Cruise (241), Zoox (167), Apple Inc (23), and Pony.ai (15), together accounting for 895 cases. The remaining 73 cases were distributed across 19 other labels, most commonly WeRide (14), Mercedes-Benz Research & Development North America, Inc (11), Woven by Toyota (11), Nuro (9), and Aurora (7). This distribution indicates that the corpus was dominated by a small number of AV reporting entities, with a comparatively small tail of less frequent manufacturers. Figure 1 shows an example of the report format used in the corpus. Each report records a traffic collision involving a permitted autonomous test vehicle, but these documents are not police generated records written at the crash scene. Under California DMV rules, manufacturers must submit a *Report of a Traffic Collision Involving an Autonomous Vehicle* (OL 316) within 10 days of any collision on a public road that results in property damage, bodily injury, or death [8, 11]. The OL 316 form is completed by the manufacturer and certified under penalty of perjury by an authorised representative or programme director, and it may also include supplementary material, such as a law enforcement report, when available [8, 9]. Accordingly, the corpus combines structured fields, including manufacturer details, vehicle information, crash time and location, injuries and damage, and environmental conditions, with a narrative account prepared for regulatory reporting rather than for on site police investigation [9, 24].

[1]For reporting consistency, legacy Google self driving and Google Auto entries were consolidated under *Waymo*, and Toyota AV software lineage entries were consolidated under *Woven by Toyota*. These merges reflect practical company lineage roll up decisions used for analysis rather than strict one to one legal entity matching.

2.2 Prompting Procedure and Response Collection

To analyse the reports, we used the ChatGPT web interface with *GPT-5.4 Thinking* and the *Standard* thinking effort (https://openai.com/index/introducing-gpt-5-4). In this study, the model was used as a controlled document analysis tool to extract structured information from visually rich collision reports under a fixed prompting protocol, rather than as a source of open ended interpretation.

The prompt design was treated as the main methodological step. Because the California autonomous vehicle collision reports are multi-page PDF documents that combine form fields, narrative text, diagrams, and checkbox-marked categories, we designed the prompts to minimise ambiguity, support consistent field extraction, and limit unsupported inference. Previous work on prompting for structured documents and document-level information extraction has similarly highlighted the importance of explicit task definition, schema-aligned outputs, and constrained response formats for improving reliability [16, 47, 49]. The instruction prompt, query document, and example reports are provided in section 5.

For each report, we initiated a new interaction and first entered an instruction prompt that defined the task and its constraints. The prompt stated that two documents would be provided: the collision report itself and a second document containing the extraction questions. It also specified that the AV should always be treated as *Vehicle 1* and the other involved road user as *Vehicle 2* or *Other Party*, consistent with the reports. To keep the extraction grounded in the source document, the prompt instructed the model to answer each question only from the designated report section, to avoid adding text beyond what was requested, and to avoid using external knowledge unless a question explicitly required online lookup. The prompt also required the model to return `NA` whenever the confidence in an answer was less than 50%. This structure was intended to support consistent role assignment, section-bounded extraction, compact field-level responses, and explicit abstention when the report did not support a confident answer [45, 47].

The query document contained 25 questions, labelled Q0–Q24, designed to extract both descriptive and analytic information from the reports. Q0 asked for an overall judgement of whether the AV was responsible for the collision, together with an explanation, a main contributing factor, and a confidence estimate. Q1–Q3 covered manufacturer details, vehicle attributes, accident date and time, and crash location. Q4–Q6 explicitly requested online lookup based on the extracted address and time to determine the number of lanes, street type, speed limit, and whether the street was likely to be busy at the reported time. Q7–Q17 extracted damage, injury, other party information, AV operating mode, movement, lane position, intersection involvement, direction of travel, and textual descriptions of damage. Here, *v1* refers to the autonomous vehicle and *v2* to the other involved party, so variables such as *v1_lane*, *v2_lane*, *v1_speed* and *v2_speed* denote the lane position and the pre-collision speed for the respective parties. Q18–Q24 focused on page 3 of the reports and targeted bounded categories such as weather, lighting, roadway surface, roadway conditions, movement preceding collision, type of collision, and other associated factors. Many of these questions referred to checkboxes or visually marked options and, therefore, required the model to recover information from the document layout and text. The complete query set and field definitions are provided in section 5.

The prompts were refined iteratively through several rounds of pilot tests before the full corpus was processed. In each round, we applied the prompt set to a randomly selected sample of reports, reviewed the output, and revised the wording of the queries to improve clarity, grounding, and consistency. This pilot phase was used to stabilise the role definitions, section-specific instructions, answer format, and abstention rule before applying the final prompt set to the full corpus.

After entering the instruction prompt, we uploaded the collision report PDF together with the query document in the same interaction. The responses generated for each report were recorded verbatim and then stored in tabular form for further parsing and analysis.

2.3 Post Processing and Derived Analytical Variables

The recorded model outputs were post processed using a rule based parsing pipeline developed for this study. The parser converted the semi structured responses to Q0–Q24 into a canonical schema by matching answer labels to predefined field aliases and normalising the extracted values. This produced structured variables covering report metadata, fault related judgments, vehicle and location information, online enrichment fields, injuries and damage, movement and intersection variables, and checkbox based environmental and collision attributes. It also derived a unified collision type from the first available collision field, standardised boolean and categorical values, and retained row level metadata together with the original raw response so that each structured entry remained traceable to the model generated text.

The pipeline then derived analysis facing variables that transformed field level extractions into grouped representations of road user type, road user vulnerability, AV mode, AV movement, other party movement, collision group, blame group, intersection context, environmental friction profile, harm scope, and scenario class. Scenario classes were assigned using explicit rule based combinations of AV movement, other party movement, collision group, intersection context, road user vulnerability, and selected cues from the explanatory text. The resulting 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*. For each assignment, the pipeline also retained the rule trigger, evidence source, and an evidence count so that later analyses could distinguish more and less determinate cases.

To characterise the quality, interpretability, and robustness of the report, we grouped the fields into four provenance types: bounded form fields, bounded checkbox fields, narrative fields, and online enriched fields. These groupings were used to calculate provenance specific availability counts and rates, report completeness, report explicitness, coarse and fine context scores, and a context granularity gap between higher level and more detailed contextual information. Report completeness was defined as the proportion of non missing values across a fixed set of core fields covering collision type, blame related fields, weather and lighting, and selected lane, speed, and intersection variables. Report explicitness was defined as a weighted availability score in which bounded form and checkbox fields contributed more strongly than narrative fields, and online enriched fields contributed the least. The coarse context score was defined as the proportion of contextual variables available at the higher level, while the fine context score was defined as the proportion of detailed contextual variables available, such as lane position, speed, intersection participation, direction of travel, lane number, street type, and street busyness. The context granularity gap was calculated as the difference between the coarse and fine context scores.

Additional diagnostic variables were derived to assess internal consistency and interpretability. These included scenario determinability, movement consistency status, stopped AV subtype, intersection detail quality, blame conflict flags, blame confidence alignment, blame evidence strength, and external enrichment level. Together, these variables allowed us to analyse not only recurring collision scenarios, but also how strongly those scenarios were supported by the reporting infrastructure.

Missing values were retained as `NA` instead of removed by complete case filtering. This allowed incomplete or under specified reports to remain part of the analysis while making patterns of missing data visible in later blind spot summaries and descriptive figures. This choice follows broader guidance in missing data analysis, which warns against routine list deletion when incomplete observations may still contain useful information and when patterns of missing data may themselves be analytically meaningful [20, 51]. In the final configuration used in this study, all 970 parsed rows were retained for analysis, and missingness was explicitly examined through blind spot summaries, contradiction audits, and related review output. The post processed dataset was then exported together with run level summary tables, including taxonomy counts, field provenance summaries, movement inconsistency audits, contradiction reports, and validation samples for manual review.

2.4 Analytical Strategy

The analysis was organised in two complementary stages. In the first stage, we used the structured dataset to characterise recurring patterns in collisions involving AVs. This stage focused on descriptive distributions of scenario class, road user type, blame attribution, environmental profile, and related grouped variables, together with visual summaries such as general Sankey diagrams, sunburst plots, categorical histograms, and taxonomy focused figures. The goal of this stage was to identify the dominant collision situations represented in the corpus and to examine how these patterns varied between road users, movement configurations, and accountability groups.

In the second stage, we examine the reporting and extraction process itself. Rather than treating all structured variables equally well supported, we analysed missingness, field provenance, completeness, explicitness, coarse versus fine context coverage, scenario determinability, movement consistency, blame confidence alignment, and contradiction signals. This stage was intended to assess not only what the reports reveal about recurring AV collision scenarios, but also how strongly these interpretations are supported by the reporting format and the extracted evidence.

Importantly, incomplete cases were retained in the analysis rather than excluded by listwise deletion. Missing values were preserved as `NA` and explicitly examined by blind spot summaries and descriptive figures. This choice follows broader guidance in missing data analysis, which warns against routine complete case filtering when incomplete observations may still contain analytically useful information and when missingness itself may be substantively informative [20, 51]. Taken together, this two stage strategy allowed us to study AV collision reports not only as a source of crash data, but also as a reporting infrastructure whose structure supports some forms of interpretation while leaving others underspecified or fragile.

3 Results

The final corpus contained 970 report entries. In the final pipeline configuration, all 970 entries had a usable model generated response in the selected *Output* column, and all 970 were parsed successfully. Figure 2 provides a corpus level overview of how cases are distributed across road user type, AV mode, movement preceding collision, collision group, blame group, and scenario class. Across the entire corpus, the mean report explicitness score was 0.79, indicating that a substantial share of report information could be recovered in structured form, although not all field types were equally well 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 the broad scenario structure much more consistently than the detailed context at the interaction level.

The scenario distribution was concentrated in a small number of recurring patterns. The largest scenario class was *AV stopped rear end* with 266 cases (27.42%), followed by *other or ambiguous* with 212 cases (21.86%), *intersection lateral conflict* with 180 cases (18.56%) and *lane change or merge conflict* with 156 cases (16.08%). The remaining scenario classes were *curbside or parked vehicle conflict* with 116 cases (11.96%), *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 overall

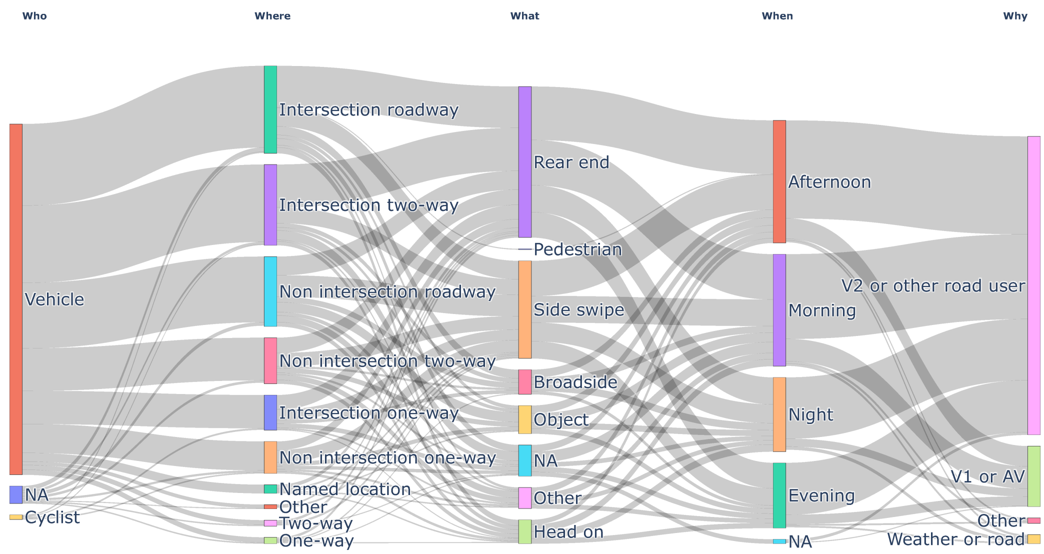

Fig. 2. Sankey diagram showing how collision cases are distributed across the main analytical dimensions in the dataset.

pattern is shown in Figure 3. The other involved party was classified as a vehicle in 901 cases (92.89%), unknown in 46 cases (4.74%), a cyclist in 14 cases (1.44%), a pedestrian in 3 cases (0.31%) and a skateboarder, a steel drain cover, a kerb, an animal, a utility valve cover, and a skateboard in 1 case each (0.10% each). The reported blame was assigned to the other road user in 804 cases (82.89%), to the AV as the primary party in 136 cases (14.02%), to environmental or road conditions in 28 cases (2.89%) and remained unclear in 2 cases (0.21%). Figure 4 shows that these accountability patterns varied across scenario classes rather than being distributed uniformly in all collision situations.

The reports were substantially more informative for coarse scenario structure than for fine grained interaction detail. The highest missingness rates were observed for *v2 speed* (87.94%), *v2 lane* (78.04%), *street busy* (77.84%), *v1 lane* (76.49%) and *v1 speed* (75.57%). The missing was also substantial for *the description of the damage of* v2 (52.47%), *the number* of lane (49.79%) and *the type* of the street (44.12%). Mean field availability differed across provenance groups, with the highest availability for bounded form fields (0.85), followed by bounded checkbox fields (0.76), narrative fields (0.70) and online enriched fields (0.48). This pattern is consistent with the difference between the mean coarse context score (0.97) and the mean fine context score (0.48), indicating that the reports captured a broad scenario structure much more consistently than the detailed context of the interaction level. Scenario determinability was classified as high in 720 cases (74.23%), medium in 241 cases (24.85%) and low in 9 cases (0.93%), indicating that most cases could be assigned to a scenario class, although often with weaker support for detailed contextual reconstruction.

Additional indicators pointed to substantial limits in interaction level detail. Movement consistency was classified as consistent in 284 cases (29.28%) and inconsistent in 686 cases (70.72%). Among inconsistent cases, 515 (75.07%) were labelled as within the report conflict or parser issue and 171 (24.93%) as context of an ambiguous scenario. The strength of the blame evidence was classified as strong in 943 cases (97.22%), conflicting in 25 cases (2.58%) and weak in 2 cases

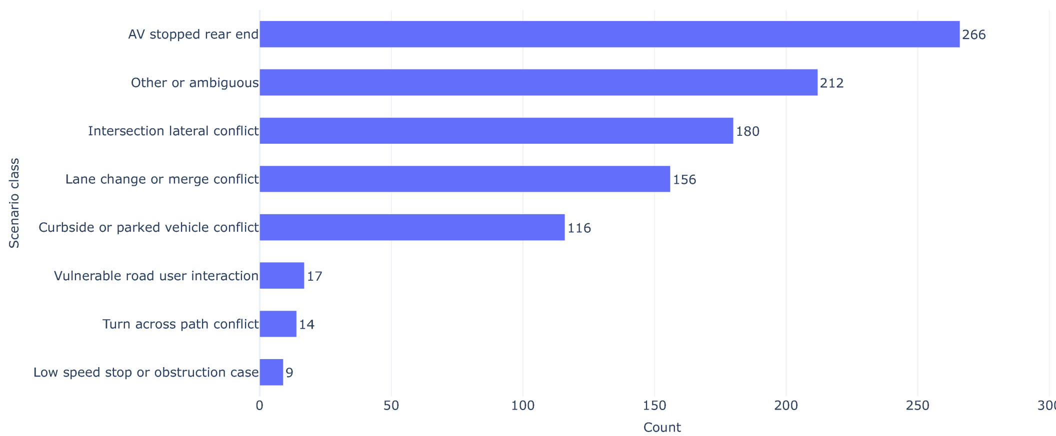

Fig. 3. The distribution of scenario classes in the collision corpus. Each bar represents the number of cases assigned to a given scenario class, allowing comparison of the most common recurring collision situations, including AV stopped rear end crashes, intersection lateral conflicts, lane change or 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.

(0.21%), while the alignment of the blame confidence was classified as aligned or unclear in 931 cases (95.98%) and as high confidence with low evidence in 39 cases (4.02%). The friction profile of the environment was classified as nominal in 860 cases (88.66%), compounded degradation in 36 cases (3.71%), visibility degraded in 33 cases (3.40%), unusual road surface in 25 cases (2.58%) and surface degraded in 16 cases (1.65%). Among the stopped AV cases, the largest subtype was *the rear end* stopped with 269 cases (56.99%), followed by *stopped at the intersection* with 108 cases (22.88%), *traffic stop* with 67 cases (14.19%), *stop for obstruction or uncertainty* with 23 cases (4.87%) and *other stopped case* with 5 cases (1.06%). The quality of the details of the intersection was classified as contextualised intersection in 502 cases (51.75%), not intersection focused in 444 cases (45.77%) and intersection flag only in 24 cases (2.47%). Taken together, these results show that the reports support broad scenario classification and blame attribution more consistently than detailed reconstruction of lane level, speed related and contextual interaction dynamics.

As shown in Figure 5, the field level missingness profile reinforces this pattern. The most severe blind spots were concentrated in lane position, speed, and traffic context, whereas higher level contextual variables such as direction of travel and intersection involvement were available for most reports. This suggests that the reports are considerably more informative for broad scenario characterisation than for fine grained reconstruction of how the interaction unfolded in detail.

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 Figure 2 and Figure 3, these scenario classes account for most of the corpus, although a substantial minority of reports remained in the *other or ambiguous* category. 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. In such

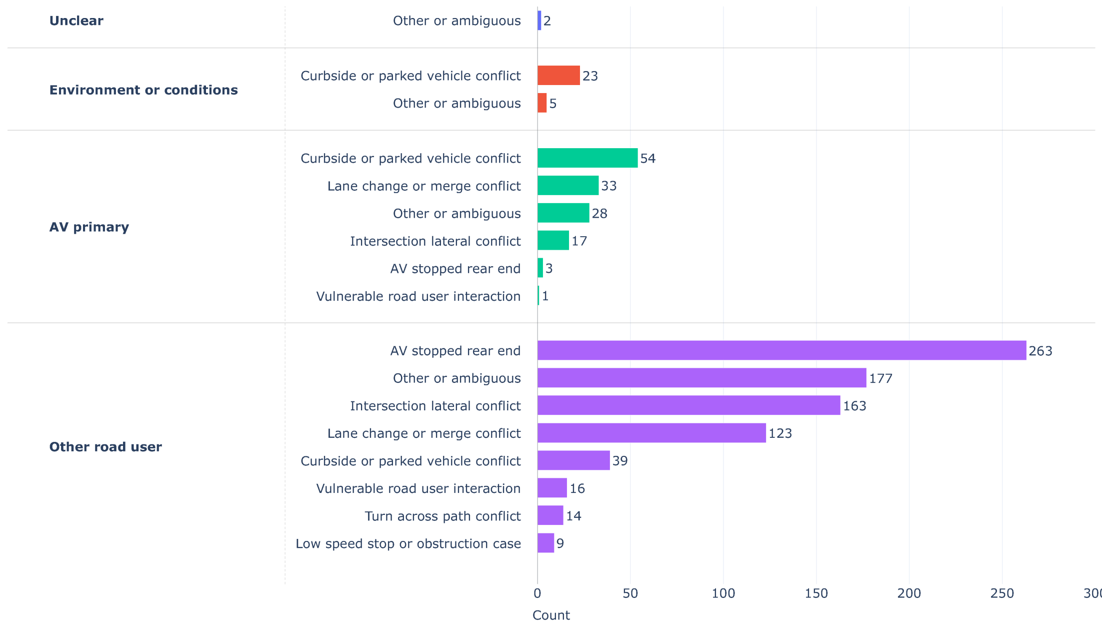

Fig. 4. The distribution of blame attribution across scenario classes in the collision corpus. For each scenario class, the bars indicate how many cases were assigned to the AV, the other road user, environmental or road condition factors, or remained unclear, allowing comparison of accountability patterns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

situations, the challenge is not only for the AV to perceive the scene correctly, but also to anticipate the behaviour of other road users and to respond in ways that align with human expectations. The prominence of rear end and other low speed interaction cases is broadly consistent with earlier studies of California DMV reports [7, 24, 25]. Our contribution extends this line of work by organising these collisions into a more explicit scenario based structure, rather than treating them primarily as separate crash variables.

The prominence of *AV stopped rear end* and *intersection lateral conflict* likely reflects the particular demands of mixed traffic interaction. At intersections, road users must judge priority, speed, intent, and possible future paths in a short time, often under uncertainty and without complete information. These situations require not only perception, but also the interpretation of behavioural cues and expectations. A similar challenge appears in stopped AV rear end crashes. In these cases, the AV may behave conservatively or legally, but the driver behind it may not expect the timing or extent of its braking or stopping. More generally, these findings suggest that safety problems in mixed traffic are not limited to dramatic system failures. They also emerge from ordinary but interaction heavy situations in which AVs and human road users must continuously coordinate with each other. The high share of cases assigned to the other road user in the reported blame distribution, shown in Figure 4, is consistent with this interpretation and with previous work on California AV crash reports [24, 59]. At the same time, this pattern should be interpreted as *reported* attribution rather than as direct ground truth about causal responsibility, because fault in these reports is itself shaped by narrative framework, reporting practice, and legal context.

A second major finding is the large gap between coarse and fine contextual information. The reports were generally sufficient for broad scenario assignment, but often lacked lane level, speed related, and traffic context detail. This pattern

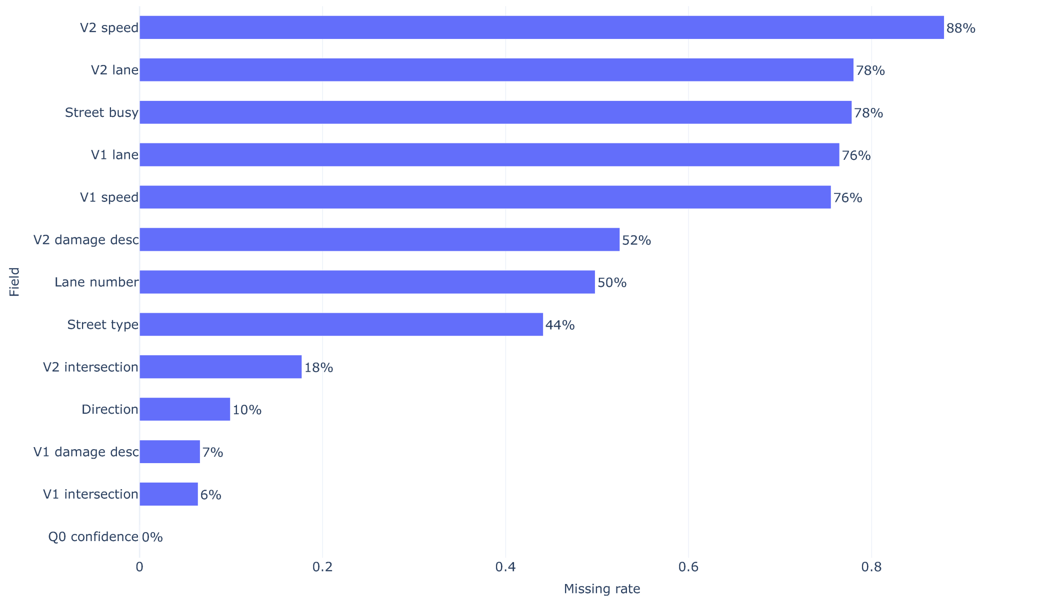

Fig. 5. Field-level missingness for selected blind-spot variables in the collision report corpus. The figure shows the proportion of reports in which fine-grained contextual variables, including speed, lane position, street busyness, lane number, street type, intersection detail, direction and damage 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.

is reflected both in the high missingness rates for variables such as *v2 speed*, *v2 lane*, *street busy*, *v1 lane* , and *v1 speed* and in the large difference between the mean coarse context score and the mean fine context score. By contrast, higher level variables such as scenario class, blame group, and collision group were much more consistently available. One plausible reason is that the California collision form is designed primarily for administrative documentation and legal reporting rather than for detailed reconstruction of the interaction. Broad facts such as crash type, location, operating mode, and damage are easier to record in standardised form, whereas fine grained movement and interaction detail often depend on narrative description, interpretation of the scene, or information that may not have been observed or documented systematically at the time of reporting. This helps explain why the reports are well suited to identify broad crash patterns but less well suited to describing the detailed structure of negotiation and movement in mixed traffic.

The provenance analysis supports this interpretation. Bounded form fields had the highest mean availability, followed by bounded checkbox fields and narrative fields, while online enriched variables had the lowest availability. This suggests that the most consistently recoverable information is information that the reporting form already expects and structures explicitly. Narrative fields remain valuable, but their usefulness depends on how the event was described, by whom, and with how much detail. Online enriched variables add potentially useful contextual information, but they also introduce an additional layer of uncertainty because they are not recorded directly in the report itself. Together, these findings indicate that the reporting infrastructure shapes the kinds of question that can be answered reliably. The reports and our interpretation pipeline are well suited to recovering the occurrence of a crash and some of its broad properties, but they are less robust for recovering the fine grained interaction context that would be needed for deeper behavioural or causal analysis.

The results of determinability and consistency reinforce this point. Most reports were classified as having high or medium scenario determinability, indicating that available information was generally sufficient for a broad scenario assignment. At the same time, the large size of the *other or ambiguous* category shows that not all reports can be confidently placed into a narrow scenario class. This may reflect genuinely unusual events, but it may also result from unspecified reports, incomplete narratives, or missing fine context. In other words, ambiguity is not only a property of the crash itself. It can also be a property of the way the crash was documented. The high level of inconsistency in the movement points in the same direction. These inconsistencies should not be interpreted simply as a model error. More likely, they reflect a combination of partial mismatch between structured and narrative fields, limits of the parser, and genuine ambiguity in the source reports themselves.

The environmental results add an additional layer of interpretation. Most reports were classified as occurring under nominal conditions, with only a small share involving degraded visibility, unusual road conditions, or degraded surface conditions. This suggests that many reported collisions did not require extreme weather or severe road conditions to occur. Instead, they occurred under conditions that were relatively ordinary from an environmental point of view. One possible explanation is that interaction complexity, rather than physical degradation, is the most common challenge in these reports. Another is that environmental conditions are easier to record in coarse categories than the detailed dynamics of the interaction that unfolds between road users. The results of the stopped AV subtype are especially consistent with this view. Most stopped AV cases were classified as stopped rear end crashes, followed by stopped at intersection and traffic stop cases. This pattern is also broadly in line with earlier observations that AVs are often struck from behind while stopped [25]. These cases are important because they sit at the boundary between lawful AV behaviour and the expectations of surrounding road users. They therefore point to the importance of understanding not only whether an AV followed the rules but also how its behaviour is interpreted by others in the traffic stream.

Taken together, the findings suggest that AV collision reports can support meaningful large scale analysis of recurring crash scenarios, but that they also reveal clear limits in what can be recovered reliably. The reports support broad scenario classification, descriptive blame patterns, and high level contextual analysis. They are less well suited for detailed reconstruction of motion, lane level interaction, and traffic context. This distinction is important both for research and for practice. For research, collision reports suggest that they are most informative when used to identify recurring interaction patterns and reporting gaps rather than to infer detailed causal chains from the sparse evidence. In practice, it suggests that current reporting forms could be improved by capturing more explicit information about lane position, relative motion, intersection configuration, and the behaviour of the other road user. Such changes would make future reports more useful not only for compliance and investigation, but also for understanding the human and interactional challenges of autonomous driving in mixed traffic.

4.1 Limitations and Future Work

The findings of this study should be interpreted in light of the source on which they are based. The analysis is based on collision reports from a single reporting system, namely the California Department of Motor Vehicles. Although this corpus is important and has been used in earlier studies of AV collisions, it reflects reporting practices, regulatory setting, operational conditions, and road use norms of only one jurisdiction [18, 22, 24, 46]. The results should therefore be read as findings from this reporting context rather than as a complete account of AV safety, more generally. In particular, the study does not support conclusions about how collision patterns, reporting practices, or interaction norms may differ between jurisdictions or cultural contexts. This matters because the reports are not ground truth records of the crash. They are secondary documents that can omit detail, simplify events, or frame responsibility in ways that do not fully capture the causal chain. In autonomous driving, especially, rare but consequential events must be interpreted against a much broader background of overall system performance [34]. For that reason, the present study characterises reported crashes and reported crash circumstances rather than the full safety performance of AVs in naturalistic driving.

A further limitation lies in the structure and completeness of the reports themselves. Several variables that would be highly informative for the analysis of interaction, including lane position, speed, and traffic context, were frequently missing or under specified. Earlier work on California AV reports has similarly noted limited narrative detail and difficulty in determining some variables directly from reports [3, 18]. As a result, the taxonomy developed here is more robust at the level of broad scenario classes than at the level of detailed movement reconstruction. This is not only a limitation of the extraction workflow, but also of the reporting infrastructure from which the evidence is drawn. Retaining incomplete cases as NA made these blind spots visible, but it did not resolve the underlying imbalance by which some aspects of interaction are documented much more sparsely than others.

The analytical pipeline also introduces its own constraints. The variables used in the analysis were derived through a combination of model generated responses and rule based post processing. This design was intended to preserve transparency and traceability, yet it also means that some results depend on modelling and coding choices made by the authors. Scenario classes, blame related indicators, determinability, and consistency measures are therefore derived representations rather than direct observations from the reports themselves. Although the prompts were refined through pilot testing and the outputs were parsed under a fixed protocol, the study does not yet provide full external validation against an independently hand coded gold standard. The online enrichment variables, including lane number, street type, speed limit, and inferred street busyness, introduce an additional source of uncertainty because they depend on external search rather than solely on the report. More broadly, reports are visually structured documents that combine checkboxes, layout cues, and form elements that are not fully reducible to plain text. Recent work on multimodal document understanding suggests that models designed for layout rich documents may offer stronger support for extracting this kind of visually encoded information [31, 37].

These limitations also point to several directions for future work. An important next step would be to validate the extracted variables in a manually annotated subset, allowing scenario class, blame group, determinability, and related outputs to be systematically compared against human coding. It would also be valuable to compare the present workflow across different models, prompt strategies, and document understanding approaches, including multimodal systems designed explicitly for forms such as inputs and visually structured pages [22, 31, 37]. The scope of the analysis could be further broadened by linking collision reports with other data sources, such as disengagement reports, police records, insurance records, or naturalistic safety data, to examine whether the scenario patterns observed here extend beyond the California DMV corpus [25, 34]. Comparative studies across jurisdictions would also be valuable for examining whether reporting practices, mixed traffic interaction patterns, and road user expectations differ across regulatory and cultural settings. There is also a clear potential to connect this kind of report analysis to scenario generation and safety tests, as recent work has shown that accident reports can be transformed into structured and reusable scenarios for the evaluation of autonomous driving [26, 53]. Finally, the findings suggest that future work should not only analyse collision reports, but also improve them. More explicit reporting fields for lane position, relative motion, road user intent, intersection configuration, and control transitions would make these reports more useful both for crash investigation and for studying human interaction with autonomous vehicles in mixed traffic [2, 19, 27, 29].

5 Supplementary Material

In line with current open science practices and recommendations for transparency in automotive research [21], 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.

References

[1] Maryam Akbari, B Lankarani Kamran, Seyed Taghi Heydari, Seyed Abbas Motevalian, Reza Tabrizi, Zohreh Asadi-Shekari, and Mark JMSullman. 2019. Meta-analysis of the correlation between personality characteristics and risky driving behaviors. *Journal of injury and violence research* 11, 2 (2019), 107. https://doi.org/10.5249/jivr.v11i2.1172

[2] 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. https://doi.org/10.1145/3744335.3758477

[3] 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).

[4] Victoria A Banks, Katherine L Plant, and Neville A Stanton. 2018. Driver error or designer error: Using the Perceptual Cycle Model to explore the circumstances surrounding the fatal Tesla crash on 7th May 2016. *Safety science* 108 (2018), 278–285. https://doi.org/10.1016/j.ssci.2017.12.023

[5] Pavlo Bazilinskyy, Dimitra Dodou, Yke Bauke Eisma, Willem Vlakveld, and Joost de Winter. 2023. Blinded windows and empty driver seats: The effects of automated vehicle characteristics on cyclists' decision-making. *IET Intelligent Transport Systems* 17, 1 (2023), 72–84. https://doi.org/10.1049/itr2.12235

[6] Andrew L Berthaume, Rachel M James, Britton E Hammit, Christina Foreman, and Christopher L Melson. 2018. Variations in driver behavior: An analysis of car-following behavior heterogeneity as a function of road type and traffic condition. *Transportation research record* 2672, 37 (2018), 31–44. https://doi.org/10.1177/0361198118798713

[7] Wayne Biever, Linda Angell, and Sean Seaman. 2020. Automated driving system collisions: Early lessons. *Human factors* 62, 2 (2020), 249–259. https://doi.org/10.1177/0018720819872034

[8] California Code of Regulations. 2024. Title 13, Section 227.48: Reporting Collisions. https://www.law.cornell.edu/regulations/california/13-CCR-227.48 Accessed 2026-04-07.

[9] California Department of Motor Vehicles. 2020. Report of Traffic Collision Involving an Autonomous Vehicle (OL 316). https://www.dmv.ca.gov/portal/file/report-of-traffic-accident-involving-an-autonomous-vehicle-ol-316-pdf/ Rev. 7/2020, accessed 2026-04-07.

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

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

[12] California Department of Motor Vehicles. 2026. Autonomous Vehicles Terms and Definitions. https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-definitions/ Accessed 2026-04-07.

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

[14] Hengrui Chen, Hong Chen, Zhizhen Liu, Xiaoke Sun, and Ruiyu Zhou. 2020. Analysis of factors affecting the severity of automated vehicle crashes using XGBoost model combining POI data. *Journal of advanced transportation* 2020, 1 (2020), 8881545. https://doi.org/10.1155/2020/8881545

[15] Amit Chougule, Vinay Chamola, Aishwarya Sam, Fei Richard Yu, and Biplab Sikdar. 2023. A comprehensive review on limitations of autonomous driving and its impact on accidents and collisions. *IEEE Open Journal of Vehicular Technology* 5 (2023), 142–161. https://doi.org/10.1109/OJVT.2023.3335180

[16] Gaye Colakoglu, Gürkan Solmaz, and Jonathan Fürst. 2025. Problem solved? information extraction design space for layout-rich documents using llms. *Findings of the Association for Computational Linguistics: EMNLP 2025* (2025), 17908–17927.

[17] Jin Cui, Lin Shen Liew, Giedre Sabaliauskaite, and Fengjun Zhou. 2019. A review on safety failures, security attacks, and available countermeasures for autonomous vehicles. *Ad Hoc Networks* 90 (2019), 101823. https://doi.org/10.1016/j.adhoc.2018.12.006

[18] 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. https://doi.org/10.1016/j.iatssr.2020.03.001

[19] 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. https://doi.org/10.1016/j.trip.2020.100174

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

[21] 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. https://doi.org/10.1145/3640792.3675730

[22] 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. https://doi.org/10.3390/electronics14040651

[23] RM Faber, O Jonkeren, MC de Haas, EJE Molin, and M Kroesen. 2022. Inferring modality styles by revealing mode choice heterogeneity in response to weather conditions. *Transportation research part A: policy and practice* 162 (2022), 282–295. https://doi.org/10.1016/j.tra.2022.06.003

[24] 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. https://doi.org/10.1371/journal.pone.0184952

[25] 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. https://doi.org/10.1016/j.aap.2021.106056

[26] An Guo, Yuan Zhou, Haoxiang Tian, Chunrong Fang, Yunjian Sun, Weisong Sun, Xinyu Gao, Anh Tuan Luu, Yang Liu, and Zhenyu Chen. 2024. SoVAR: Build Generalizable Scenarios from Accident Reports for Autonomous Driving Testing. In *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering* (Sacramento, CA, USA) *(ASE '24)*. Association for Computing Machinery, New York, NY, USA, 268–280. https://doi.org/10.1145/3691620.3695037

[27] Jiawen Guo, Quan Yuan, Jingrui Yu, Xizheng Chen, Wenlin Yu, Qian Cheng, Wuhong Wang, Wenhui Luo, and Xiaobei Jiang. 2022. External human–machine interfaces for autonomous vehicles from pedestrians' perspective: A survey study. *Sensors* 22, 9 (2022), 3339. https://doi.org/10.3390/s22093339

[28] Azra Habibovic, Victor Malmsten Lundgren, Jonas Andersson, Maria Klingegård, Tobias Lagström, Anna Sirkka, Johan Fagerlönn, Claes Edgren, Rikard Fredriksson, Stas Krupenia, et al. 2018. Communicating intent of automated vehicles to pedestrians. *Frontiers in psychology* 9 (2018), 284756. https://doi.org/10.3389/fpsyg.2018.01336

[29] 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. https://doi.org/10.1080/1463922X.2019.1574931

[30] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434. https://doi.org/10.1177/0018720814547570

[31] Yupan Huang, Tengchao Lv, Lei Cui, Yutong Lu, and Furu Wei. 2022. Layoutlmv3: Pre-training for document ai with unified text and image masking. In *Proceedings of the 30th ACM international conference on multimedia*. 4083–4091. https://doi.org/10.1145/3503161.3548112

[32] Rachel M James and Britton E Hammit. 2019. Identifying contributory factors to heterogeneity in driving behavior: clustering and classification approach. *Transportation research record* 2673, 10 (2019), 343–353. https://doi.org/10.1177/0361198119849404

[33] Satish Jeyachandran. 2024. *Meet the 6th-generation Waymo Driver: Optimized for costs, designed to handle more weather, and coming to riders faster than before.* https://waymo.com/blog/2024/08/meet-the-6th-generation-waymo-driver/ Waymo Blog, 19 August 2024. Accessed on 2 April 2026.

[34] Nidhi Kalra and Susan M Paddock. 2016. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation research part A: policy and practice* 94 (2016), 182–193. https://doi.org/10.1016/j.tra.2016.09.010

[35] Geewook Kim, Teakgyu Hong, Moonbin Yim, JeongYeon Nam, Jinyoung Park, Jinyeong Yim, Wonseok Hwang, Sangdoo Yun, Dongyoon Han, and Seunghyun Park. 2022. Ocr-free document understanding transformer. In *European Conference on Computer Vision*. Springer, 498–517. https://doi.org/10.1007/978-3-031-19815-1_29

[36] Moritz Körber, Eva Baseler, and Klaus Bengler. 2018. Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied ergonomics* 66 (2018), 18–31. https://doi.org/10.1016/j.apergo.2017.07.006

[37] Kenton Lee, Mandar Joshi, Iulia Raluca Turc, Hexiang Hu, Fangyu Liu, Julian Martin Eisenschlos, Urvashi Khandelwal, Peter Shaw, Ming-Wei Chang, and Kristina Toutanova. 2023. Pix2struct: Screenshot parsing as pretraining for visual language understanding. In *International Conference on Machine Learning*. PMLR, 18893–18912. https://doi.org/10.5555/3618408.3619188

[38] Shervin Hajinia Leilabadi and Stephan Schmidt. 2019. In-depth analysis of autonomous vehicle collisions in California. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 889–893. https://doi.org/10.1109/ITSC.2019.8916775

[39] Carl Macrae. 2022. Learning from the failure of autonomous and intelligent systems: Accidents, safety, and sociotechnical sources of risk. *Risk analysis* 42, 9 (2022), 1999–2025. https://doi.org/10.1111/risa.13850

[40] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. 2018. Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12. https://doi.org/10.1145/3173574.3174003

[41] Gerald Matthews, Jinchao Lin, April Rose Panganiban, and Michael D Long. 2019. Individual differences in trust in autonomous robots: Implications for transparency. *IEEE Transactions on Human-Machine Systems* 50, 3 (2019), 234–244. https://doi.org/10.1109/THMS.2019.2947592

[42] 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).

[43] National Transportation Safety Board. 2017. *Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida, May 7, 2016.* Highway Accident Report NTSB/HAR-17/02. National Transportation Safety Board, Washington, DC. https://www.ntsb.gov/investigations/accidentreports/reports/har1702.pdf

[44] National Transportation Safety Board. 2019. *Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, Tempe, Arizona, March 18, 2018.* Highway Accident Report NTSB/HAR-19/03. National Transportation Safety Board, Washington, DC. https://www.ntsb.gov/investigations/accidentreports/reports/har1903.pdf

[45] OpenAI. 2026. Prompt guidance for GPT-5.4. https://developers.openai.com/api/docs/guides/prompt-guidance. Accessed 2026-04-04.

[46] 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. https://doi.org/10.1177/1087724X231217677

[47] Chengfeng Qiu, Lifeng Zhou, Kaifeng Wei, and Yuke Li. 2025. DocIE@ XLLM25: UIEPrompter: A Unified Training-Free Framework for universal document-level information extraction via Structured Prompt. In *Proceedings of the 1st Joint Workshop on Large Language Models and Structure Modeling (XLLM 2025)*. 317–321.

[48] 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. https://doi.org/10.1109/TITS.2019.2901817

[49] Jon Saad-Falcon, Joe Barrow, Alexa Siu, Ani Nenkova, Seunghyun Yoon, Ryan A Rossi, and Franck Dernoncourt. 2024. Pdftriage: Question answering over long, structured documents. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*. 153–169.

[50] Dorsa Sadigh, Katherine Driggs-Campbell, Alberto Puggelli, Wenchao Li, Victor Shia, Ruzena Bajcsy, Alberto L Sangiovanni-Vincentelli, S Shankar Sastry, and Sanjit A Seshia. 2014. Data-driven probabilistic modeling and verification of human driver behavior. In *AAAI Spring Symposium-Technical Report*. 56–61. https://cdn.aaai.org/ocs/7749/7749-34383-1-PB.pdf

[51] Joseph L Schafer and John W Graham. 2002. Missing data: our view of the state of the art. *Psychological methods* 7, 2 (2002), 147.

[52] Taniya Sultana and Hany M Hassan. 2024. Does recognizability of connected and automated vehicles (CAVs) platoons affect drivers' behavior and safety? *Transportation research part F: traffic psychology and behaviour* 103 (2024), 368–386. https://doi.org/10.1016/j.trf.2024.04.021

[53] Shuncheng Tang, Zhenya Zhang, Jixiang Zhou, Lei Lei, Yuan Zhou, and Yinxing Xue. 2024. LeGEND: A Top-Down Approach to Scenario Generation of Autonomous Driving Systems Assisted by Large Language Models. In *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering* (Sacramento, CA, USA) *(ASE '24)*. Association for Computing Machinery, New York, NY, USA, 1497–1508. https://doi.org/10.1145/3691620.3695520

[54] Tesla, Inc. n.d.. *Model Y Owner's Manual: Cameras*. https://www.tesla.com/ownersmanual/modely/en_us/GUID-682FF4A7-D083-4C95-925A-5EE3752F4865.html Tesla official documentation. Accessed: 2026-04-02.

[55] Chantal PM Timmermans, Wael KM Alhajyaseen, Veerle Ross, and Hideki Nakamura. 2020. Introducing a multi-variate classification method: Risky driving acceptance among different heterogeneous driver sub-cultures. *Journal of safety research* 73 (2020), 81–91. https://doi.org/10.1016/j.jsr.2020.02.009

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

[57] Jorge Vargas, Suleiman Alsweiss, Onur Toker, Rahul Razdan, and Joshua Santos. 2021. An overview of autonomous vehicles sensors and their vulnerability to weather conditions. *Sensors* 21, 16 (2021), 5397. https://doi.org/10.3390/s21165397

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

[59] Song Wang and Zhixia Li. 2019. Exploring the mechanism of crashes with automated vehicles using statistical modeling approaches. *PloS one* 14, 3 (2019), e0214550. https://doi.org/10.1371/journal.pone.0214550

[60] Jahnavi Yarlagadda and Digvijay S Pawar. 2022. Heterogeneity in the Driver Behavior: An Exploratory Study Using Real-Time Driving Data. *Journal of advanced transportation* 2022, 1 (2022), 4509071. https://doi.org/10.1155/2022/4509071

[61] 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. https://doi.org/10.1109/ACCESS.2020.2983149

[62] Meng Zhang, Mandy Dotzauer, and Caroline Schießl. 2022. Analysis of implicit communication of motorists and cyclists in intersection using video and trajectory data. *Frontiers in psychology* 13 (2022), 864488. https://doi.org/10.3389/fpsyg.2022.864488