SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
http://saucis.sakarya.edu.tr/
e-ISSN: 2636-8129
Publisher: Sakarya University
Vol. 8, No. 4, 718-739, 2025
DOI: 10.35377/saucis...1666618

RESEARCH ARTICLE

Enhancing Autonomous Vehicle Safety Through Chaid Modeling: Influential Factors, Seasonal Variations, and Systematic Limitations

Ferhan Baş Kaman[1*], Ahmet Yücel[2]

[1] Department of Finance and Banking, Şereflikoçhisar Faculty of Applied Science, Ankara Yıldırım Beyazıt University, Ankara, Türkiye. E-mail: ferhanbaskaman@aybu.edu.tr
[2]Department of Finance and Banking, Şereflikoçhisar Faculty of Applied Science, Ankara Yıldırım Beyazıt University, Ankara, Türkiye. E-mail: ahmetyucel@aybu.edu.tr

Corresponding author:
Ferhan Baş Kaman, Ankara Yıldırım Beyazıt University,
Ankara, Türkiye,
ferhanbaskaman@aybu.edu.tr
ror.org/05ryemn72

Article History:
Received: 27.03.2025
Revised: 13.10.2025
Accepted: 16.10.2025
Published Online: 11.12.2025

ABSTRACT
The growing prominence of artificial intelligence has driven transformative innovations across sectors, with autonomous vehicles representing a salient manifestation of this technological shift. The reliability of autonomous vehicles plays a crucial role in determining their societal acceptance and large-scale deployment. Within this context, disengagement data serve as an objective indicator of system reliability. A rigorous analysis of disengagement data is essential for evaluating the real-world performance and operational reliability of autonomous vehicles. Such data circumstances necessitate human intervention, thereby revealing system vulnerabilities and opportunities for improvement. Consequently, precise and transparent disengagement analyses are vital for advancing AV technology and strengthening safety. This study investigates the determinants of disengagements and contrasts human-initiated with system-initiated events. Drawing on 17,406 reports (2021–2023), CHAID models identified key triggers including environmental context, system limitations, and operational conditions. The study identified key determinants, including planning inconsistencies, detection failures, and hardware malfunctions, and revealed clear seasonal variations, with disengagements peaking in summer and autumn and declining in winter and spring. Validated CHAID models demonstrated high accuracy, underscoring the importance of comprehensive training and testing across diverse conditions to enhance effectiveness and safety.

Keywords: Reliability, Autonomous Vehicles, Quantitative Decision-Making Methods, Artificial Intelligence

1. Introduction

Autonomous vehicles (AV) constitute a rapidly expanding research domain due to their potential to transform transportation systems. Potential benefits include reduced human error, fewer traffic accidents, and improved network efficiency. However, widespread adoption hinges on demonstrable reliability and safety. As AV technology advances, robustness, accuracy, and reliability remain critical, especially in safety-sensitive contexts. Consequently, improving AV reliability has become a prerequisite for their successful integration into modern transportation ecosystems.

To conceptualize levels of autonomy and reliability, the Society of Automotive Engineers (SAE) introduced a classification system that defines six levels of autonomy, ranging from Level 0 (no automation) to Level 5 (full automation). Levels 0 to 2 assist human drivers, while Levels 3 to 5 involve fully automated systems where the vehicle monitors and controls its environment [1]. Many researchers have approached AV reliability from a multidisciplinary perspective, assessing various factors influencing performance [2–3]. Kalra and Paddock [4] emphasized the necessity of long-distance testing to evaluate AV reliability, highlighting the need for extensive data collection over prolonged periods. In response to these concerns, the California Department of Motor Vehicles (CA DMV) launched an Autonomous Vehicle Testing Program in 2014, requiring manufacturers to report disengagements and incidents annually. Disengagements, defined as instances where the autonomous system transfers control to a human driver due to system failure or safety concerns [5], are critical for assessing AV performance. Analyzing disengagement data allows researchers to pinpoint system limitations and improvement opportunities, generating insights that benefit both developers and regulators.

Cite as: F. Baş Kaman and A. Yücel, "Enhancing autonomous vehicle safety through chaid modeling: Influential factors, seasonal variations, and systematic limitations" *Sakarya University Journal of Computer and Information Sciences*, vol. 8, no. 4, pp. 718-739, 2025. doi: 10.35377/saucis...1666618

Numerous studies have utilized DMV disengagement data, but their scope has varied significantly.

Early work examined frequencies, causes, and reaction times [6-8], while subsequent research adopted statistical and machine learning methods [9-13]. Banerjee et al. [11] analyzed 5,324 reports from 2014 to 2017 using parsing, normalization, and NLP, concluding that even manufacturers with millions of test miles, such as Waymo, remained in a "burn-in" stage requiring continuous driver vigilance. Khattak et al. [12] investigated 1,758 filtered disengagements from 2014 to 2018 through statistical modeling, finding strong links between disengagements and accident occurrence, with both AV system errors and driver decision-making playing critical roles. Zhang et al. [13] processed 14,282 events from 2014 to 2020 using a scalable NLP-driven pipeline, uncovering seasonal testing patterns and highlighting common failure categories in perception, localization, and control. More recently, Baş Kaman and Olmuş [14] reviewed statistical methodologies in AV reliability, emphasizing disengagement instances and driver response durations.

Importantly, the CA DMV initially released only aggregated disengagement counts but expanded its reporting structure in November 2017 to include detailed contextual information such as location, environmental conditions, and system status. Since 2020, reporting formats have become fully standardized, and the number of reports has grown substantially due to intensified testing. These developments have opened new opportunities for comprehensive and comparable analyses.

Unlike earlier studies that were limited to pre-2020 data or smaller, less standardized samples, the present research examines 17,406 disengagement reports from 2021 to 2023, a period characterized by uniform reporting and rapid AV development. By combining natural language processing (NLP) with CHAID statistical modeling, this study moves beyond descriptive analyses to uncover latent patterns in textual disengagement descriptions and systematically evaluate predictive factors. This dual approach enables a clear distinction between system-initiated and driver-initiated disengagements, providing deeper insights into system limitations, environmental influences, and operational contexts. The findings highlight critical triggers, including planning issues, perception errors, and hardware malfunctions, while also capturing seasonal effects. This addresses a significant gap in the literature and offers new insights into the evolving reliability and safety of autonomous vehicles.

2. Data

This study analyzes a dataset comprising 17,406 AV disengagement reports collected from 2021 to 2023. These reports contain several key variables, including the AV's operational capability without a driver, the disengagement initiator (AV system or test driver), the presence of a driver, the manufacturer, permit number, the disengagement cause, date, vehicle identification number (VIN), and location. Table 1 provides insights into disengagement reports, including the number of test vehicles used and the average number of disengagements per vehicle. Table 2 details the distribution of disengagements across manufacturers and years, categorized by initiator. Table 3 presents disengagement rates per mile across different manufacturers. The analysis of disengagement data is critical for evaluating AV reliability. Lower disengagement rates indicate improvements in autonomous driving performance, while higher rates highlight areas requiring refinement. A detailed examination of technical and environmental factors affecting disengagements will enhance industry insights. Furthermore, comparing performance metrics across manufacturers will support the identification of best practices and foster advancements in AV technology.

Table 1: Total Number of Disengagement Reports | Number of Vehicles Used | Average Number of Disengagement Reports Per Vehicle by Years and Manufacturers

Manufacturer	2021			2022			2023			All Years		
APPLE INC	714	25	28.56	6295	27	233.14	2830	53	53.39	9839	57	172.61
AIMOTIVE INC	118	2	59.00	689	2	344.50	687	2	343.50	1494	3	498.00
GHOST AUTONOMY INC	0	0	------	499	9	54.44	983	9	109.22	1482	10	148.20
BOSCH	0	0	------	5	1	5.00	309	2	154.50	314	3	104.67
EASYMILE	222	1	222.00	0	0	------	0	0	------	222	1	222.00
IMAGRY INC	71	1	71.00	204	1	204.00	124	2	62.00	399	3	133.00
MERCEDES-BENZ & DNA	274	8	34.25	36	10	3.60	0	0	------	310	17	18.24
MOTIONAL AD INC	0	0	------	179	9	19.88	549	9	61.00	728	15	48.53
QUALCOMM TECHNOLOGIES INC	143	3	47.67	154	3	51.33	171	5	34.20	468	5	93.60
TOYOTA RESEARCH INSTITUTE	444	4	111.00	87	4	21.75	0	0	------	531	4	132.75
VALEO NORTH AMERICA INC	204	2	102.00	71	1	71.00	21	1	21.00	296	2	148.00
WAYMO LLC	298	153	1.95	173	115	1.50	203	135	1.50	674	311	2.17
OTHERS	292	94	3.11	240	66	3.64	117	24	4.88	649	160	4.06
All Groups	2780	293	9.49	8632	248	34.81	5994	242	24.77	17406	591	29.45

Table 2: Distribution of the Number of Reports by Manufacturer, Years and Disengagement Initiator

Disengagement Initiator	Manufacturer	2021	2022	2023	All Years
Test Driver	APPLE INC	593	5376	2175	8144
	AIMOTIVE INC	118	689	687	1494
	GHOST AUTONOMY INC	0	409	930	1339
	BOSCH	0	4	155	159
	EASYMILE	151	0	0	151
	IMAGRY INC	71	204	124	399
	MERCEDES-BENZ & DNA	165	33	0	198
	MOTIONAL AD INC	0	179	549	728
	QUALCOMM TECHNOLOGIES INC	143	154	171	468
	TOYOTA RESEARCH INSTITUTE	444	87	0	531
	VALEO NORTH AMERICA INC	199	54	20	273
	WAYMO LLC	145	167	182	494
	OTHERS	278	213	67	558
	All Groups	2307	7569	5060	14936
AV System	APPLE INC	121	919	655	1695
	AIMOTIVE INC	0	0	0	0
	GHOST AUTONOMY INC	0	90	53	143
	BOSCH	0	1	154	155
	EASYMILE	71	0	0	71
	IMAGRY INC	0	0	0	0
	MERCEDES-BENZ & DNA	109	3	0	112
	MOTIONAL AD INC	0	0	0	0
	QUALCOMM TECHNOLOGIES INC	0	0	0	0
	TOYOTA RESEARCH INSTITUTE	0	0	0	0
	VALEO NORTH AMERICA INC	5	17	1	23
	WAYMO LLC	153	6	21	180
	OTHERS	14	27	50	91
	All Groups	473	1063	934	2470

Table 3. Disengagement rates per mile for test vehicles of manufacturers by year

	2021	2022	2023
APPLE INC	0,0500	0,0478	0,0071
AIMOTIVE INC	0,0356	0,0447	0,0476
GHOST AUTONOMY INC	--------	0,0308	0,0196
BOSCH	--------	0,1806	0,4135
EASYMILE	0,6938	--------	--------
IMAGRY INC	0,0971	0,3138	0,1714
MERCEDES-BENZ & DNA	0,0046	0,0007	0,0000
MOTIONAL AD INC	0,0000	0,0273	0,0366
QUALCOMM TECHNOLOGIES INC	0,0875	0,0424	0,0513
TOYOTA RESEARCH INSTITUTE	0,0300	0,0235	--------
VALEO NORTH AMERICA INC	0,6101	0,2094	0,2019
WAYMO LLC	0,0001	0,0001	0,0001

3. Methodology

3.1. Advanced Text Processing and Semantic Analysis Workflow

Cleaning and Standardization: The raw disengagement reports obtained from multiple manufacturers exhibited format inconsistencies across years and report types. To ensure structural uniformity, Optical Character Recognition (OCR) tools (OpenCV and Tesseract) were employed to extract text from PDF reports, which were then converted into CSV format. Each record was standardized by aligning field names and temporal variables (e.g., manufacturer, year, location). Missing or unreadable entries were encoded as 'N/A' to maintain database integrity.

Filtering and Quality Control: To enhance analytical validity, incomplete and repetitive records were removed. Reports containing noninformative textual descriptions (fewer than five words) or repeated manufacturer templates were excluded. After this filtering stage, the resulting dataset ensured balanced coverage of both system-initiated and driver-initiated disengagement events across years and manufacturers.

Text Preprocessing and Tokenization: The textual field describing the "facts causing disengagement" was preprocessed through a systematic NLP pipeline—this involved lowercasing, typo correction, punctuation and stop-word removal, tokenization, and stemming. Contextually meaningful uni-grams and higher-order n-grams (bi-, tri-, and quad-grams) were extracted using binary term weighting to capture recurrent linguistic patterns associated with disengagement contexts.

Feature Engineering and Semantic Structuring: Two text-derived analytical variables were created: Named Entity Recognition (NER) and Main Cause. NER extraction identified entities such as component types, environmental conditions, and human actions. Then, Latent Semantic Analysis (LSA) was combined with Singular Value Decomposition (SVD) to detect hidden semantic relationships among these entities. SVD decomposed the term document matrix to identify principal components, allowing semantically related expressions to be grouped into conceptual clusters. These clusters were consolidated under the Main Cause variable.

Data Consolidation: After cleaning, filtering, preprocessing, and semantic modeling, the final structured dataset contained both categorical variables (e.g., Manufacturer, Year, Location, Initiator) and derived textual variables (e.g., NER, Main Cause). This consolidated dataset served as the foundation for CHAID modeling and subsequent statistical interpretation.

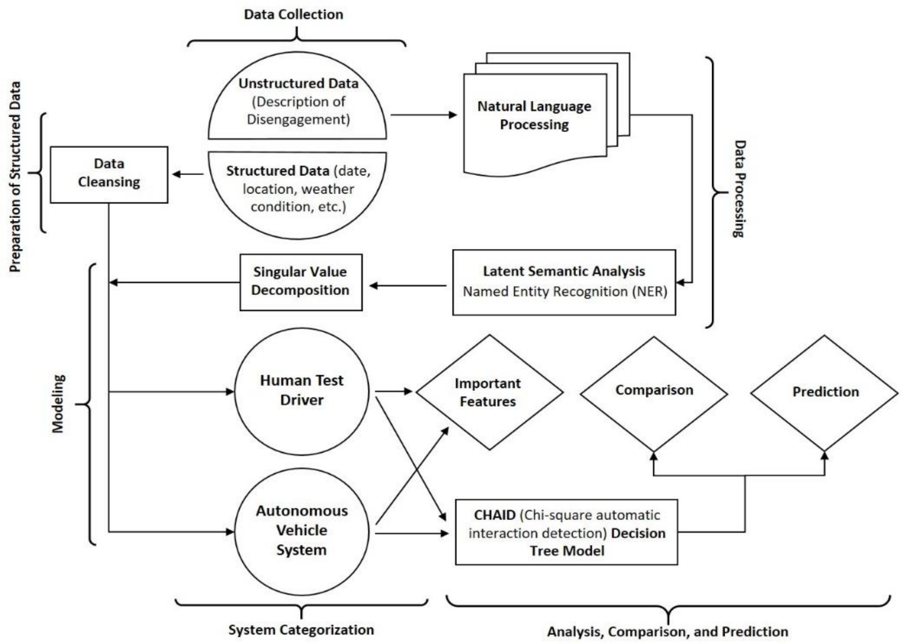

Figure 1. Methodology Diagram

To mathematically represent the semantic structure of the processed text, the term-document frequency matrix $A_{m\times n}$ was decomposed using SVD as follows:

$$A_{m\times n} = U_{m\times r} \cdot D_{r\times r} \cdot V_{r\times n}^{T} \qquad (1)$$

where $U_{m\times r}$ and $V_{n\times r}$ are orthogonal matrices and $D_{r\times r}$ contains singular values.

The diagonal elements $D_{r\times r}$ act as proportional indicators of term significance within the dataset [15].

This transformation enabled dimensionality reduction while preserving the semantic relationships necessary for subsequent CHAID-based causal modeling.

3.2. CHAID Model and Performance Evaluation in Autonomous Vehicle Disengagement Analysis

The CHAID model is a decision tree method developed by Kass [16]. This model branches the dataset using categorical data and combines explanatory variables until it finds significant differences using the chi-square (χ^2) test. CHAID offers several advantages, including its effectiveness with categorical data, ease of visualizing results, and robustness to outliers. To evaluate whether the identified factors could reliably predict disengagement initiators, a predictive decision tree model based on the CHAID algorithm was developed. This model utilized variables such as 'manufacturer', 'location', 'year', 'season', as well as the 'main cause' and 'NER' variables derived from textual data in preceding stages. Figure 1 illustrates the sequential stages of the methodology corresponding to the methods applied in this study. To ensure the robustness of the CHAID model, V-fold cross-validation was utilized for performance evaluation. V-fold cross-validation is a standard technique for assessing the performance of predictive models while minimizing bias from random data partitioning [17]. In this approach, the dataset is divided into V equally sized, non-overlapping subsets. Each subset is used once as a test set, while the remaining V-1 subsets serve as the training set. This process is iteratively repeated V times to ensure a robust evaluation of the model's performance [18]. Model performance was assessed using standard classification evaluation metrics, with a primary focus on accuracy, sensitivity, and specificity. Confusion matrices were employed to compare actual and predicted disengagement initiators, particularly in the binary classification of 'AV System' vs. 'Test Driver'. Accuracy, the most frequently used metric in unstructured data analysis, was computed using the following formula:

$$Accuracy\ rate = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

where:

True Positive (TP): Correctly classified 'AV System' cases.

True Negative (TN): Correctly classified 'Test Driver' cases.

False Positive (FP): Incorrectly classified 'AV System' cases that were actually 'Test Driver'.

False Negative (FN): Incorrectly classified 'Test Driver' cases that were actually 'AV System'.

A high accuracy rate reflects robust model performance and reliable predictive capacity. By leveraging these evaluation techniques, the reliability of classification models in AV disengagement analysis can be systematically assessed [15]. All text mining and CHAID modeling procedures were implemented in Python 3.8 using Statistica CHAID and NLP libraries (spaCy, NLTK, scikit-learn). Data preprocessing followed the same standardized steps described above.

4. Results

The 'disengagement initiated by' categorical variable in the dataset comprises two subcategories: 'Test driver' and 'AV system.' Using this variable, the data was divided into two subsets. Subsequently, word clouds were generated using the textual variable 'description of facts causing disengagement,' providing a preliminary insight into the general distribution of terms within each category. Figure 2 presents word clouds that visualize term relevance through variations in color intensity and size, providing an intuitive overview of semantic prominence. The proximity of terms in the word clouds indicates the frequency of co-occurrence within the data, providing a general understanding of the concepts influencing disengagement decisions made by the AV system and the human test driver.

Test driver decisions were primarily influenced by variables related to vehicle dynamics, lane recognition, safety control, and planning conditions. In particular, 'vehicle' and 'driver' are the two most critical factors. Words such as inaccurate, light, detection, perception, and control frequently co-occurred with driver, reflecting manual intervention triggers. Similarly, terms such as 'discrepancy', 'incorrect', 'lane', 'braking', 'stop', 'pressed', 'weather', etc., are observed to be strongly related to the term 'vehicle'. The primary factors influencing AV system decisions include concepts such as 'module', 'system', 'software', 'perception', 'planner', 'failure', etc. Particularly, 'module' and 'system' are the two most critical factors. Terms like 'error', 'performance', 'highway', 'planning', etc., appear to be strongly associated with the term 'module'. Similarly, terms such as 'data', 'component', 'trajectory', 'time', 'change', etc., are observed to be strongly related to the term 'system'.

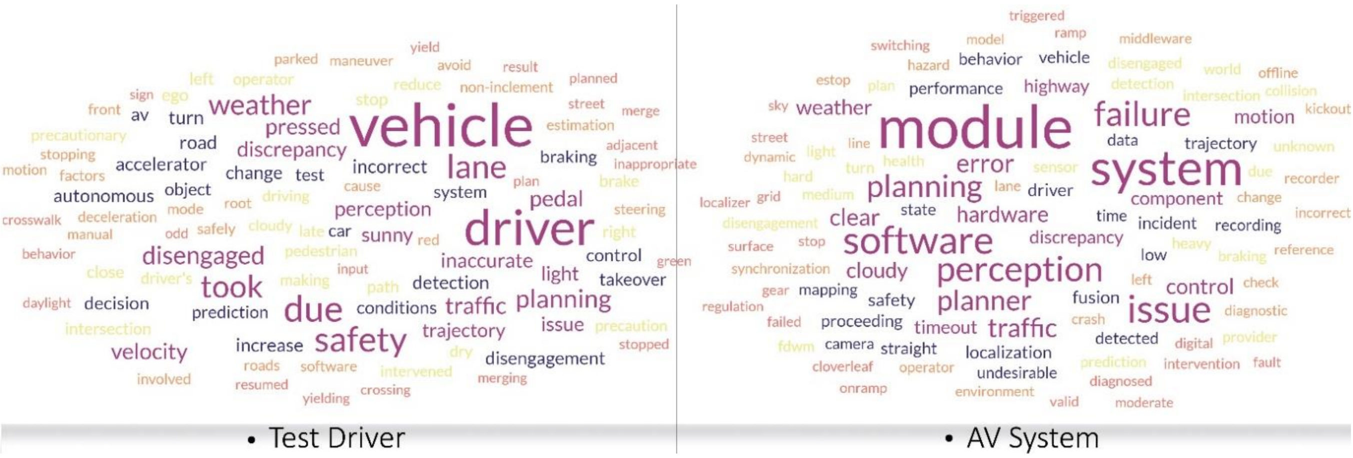

Figure 2. Word clouds

The insight presented in Figure 2 serves as an analytical framework for text analysts during the dictionary creation stage. This framework offers a comprehensive schema for identifying key terms, contextual usages, and semantic relationships. It facilitates the accurate categorization and definition of terms, ensuring that the dictionary is comprehensive, precise, and adaptable to the evolving nature of language.

The initial data was divided into two subsets based on the subcategories ('AV system' and 'Test driver') of the 'disengagement initiated by' variable. Through the application of NLP, LSA, and NER text processing techniques as summarized above, a term-document frequency table was created using binary weighting. Subsequently, the SVD method was applied to identify the significant concepts related to 'main cause' and , 'NER' in the AV system and test driver disengagement reports. The 'main cause' variable consolidates numerous factors leading to disengagement under a common term, thereby preventing the confusion that could arise from thousands of distinct concepts. Similarly, the NER variable involves the identification and classification of proper nouns and other significant entities within the text, considering only frequency values in its creation.

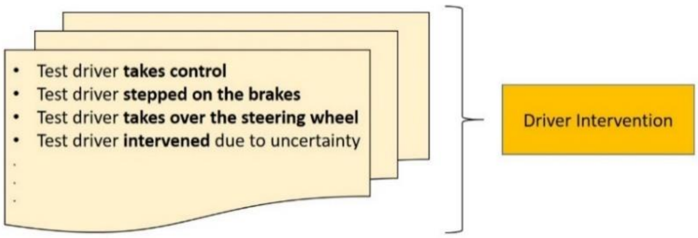

Figure 3. Determination of subcategories of the 'Main Cause' variable

Table 4. CHAID models (AV System)

Dependent (Main Cause)	Size of Node	N in class 0	N in class 1	Split (NER)	Child Node 1	Child Node 2
Driver Intervention	2470	2443	27	hardware issue	2	3
	1188	1161	27	motion plan issue	4	5
	869	842	27	software discrepancy	6	7
	721	694	27	undesirable performance	8	9
Hardware	2470	1149	1321	hardware issue	2	3
	1188	1147	41	motion plan issue	4	5
hazard detect	2470	2460	10	hardware issue	2	3
Localization	2470	2443	27	hardware issue	2	3
	1188	1161	27	motion plan issue	4	5
	869	842	27	software discrepency	6	7
	721	694	27	undesirable performance	8	9
Mapping	2470	2459	11	hardware issue	2	3
	1188	1177	11	motion plan issue	4	5
Perception	2470	2407	63	object	2	3

Table 4(Continued)

Plan	2470	2067	403	motion plan issue	2	3
	2151	2067	84	planning trajectory	4	5
	2110	2067	43	planning behavior	6	7
	2071	2067	4	planning module	8	9
Safety	2470	2460	10	hardware issue	2	3
Sensor	2470	2456	14	hardware issue	2	3
	1188	1176	12	motion plan issue	4	5
Software	2470	2386	84	hardware issue	2	3
	1188	1104	84	motion plan issue	4	5
	869	785	84	software discrepency	6	7
	721	637	84	undesirable performance	8	9
	613	529	84	object	10	11
	552	468	84	planning module	12	13
	501	417	84	planning trajectory	14	15
	460	376	84	planning behavior	16	17
System	2470	1973	497	hardware issue	2	3
	1188	691	497	motion plan issue	4	5
	869	372	497	software discrepency	6	7
	721	372	349	undesirable performance	8	9
	613	372	241	planning module	10	11
	562	368	194	object	12	13
	501	307	194	planning trajectory	14	15
	460	266	194	planning behavior	16	17

When constructing the dictionary for the 'main cause' variable, the text analyst manually aggregates concepts with different patterns but identical meanings under a single term. For example, as illustrated in Figure 3, phrases such as 'takes control,' 'stepped on the brakes,' 'takes over the steering wheel,' and 'intervened' can be grouped under the 'driver intervention' category. Due to the reports' use of varied language to express the same intent, identifying different phrases with identical meanings poses a significant challenge. Consequently, this task cannot be effectively performed using NER alone.

CHAID models were employed to elucidate the relationship between the important concepts 'main cause' and 'NER' extracted in the previous stage. The CHAID models based on AV system disengagement reports are presented in Table 4, while those based on test driver disengagement reports are shown in Table 5.

In the CHAID models, the dependent variables are binary variables derived from the subcategories of the 'main cause' variable. Similarly, the split variables in the CHAID models are binary variables created from the subcategories of the 'NER' variable. Although all split variables are significant for the corresponding dependent variable, the order of these variables in the table is presented from the most to the least important.

Table 4 provides details of the CHAID models developed for reports categorized under AV system cases. Each model describes specific scenarios that may necessitate driver intervention and identifies associated factors such as hardware problems, action plan issues, software incompatibilities, or undesirable performance. Figure 4 presents the alignment between the 'main cause' and 'NER' variables within the CHAID models.

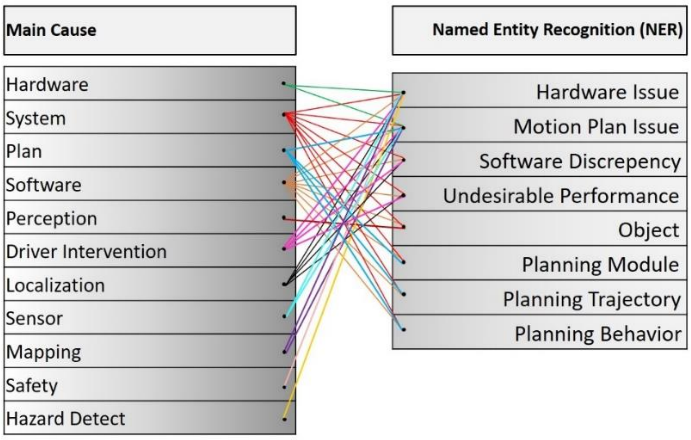

Figure 4. (AV System) Interaction between the main cause and Named Entity Recognition (NER) terms

Table 5 provides details of the CHAID models developed for reports categorized under test driver cases.

Table 5. CHAID models (Test Driver)

Dependent (Main Cause)	Size of Node	N in class 0	N in class 1	Split (NER)	Child Node 1	Child Node 2
detection	14936	14610	326	object detection	2	3
driver intervention	14936	13874	1062	precautionary disengagement	2	3
	13022	12772	250	prediction discrepency	4	5
environmental factor	14936	14662	274	prediction discrepency	2	3
	11479	11205	274	planning discrepency	4	5
	9464	9190	274	perception discrepency	6	7
	7972	7698	274	map discrepency	8	9
	6726	6452	274	sub-system	10	11
	6113	5839	274	object detection	12	13
	5773	5499	274	lane change	14	15
	5436	5162	274	safety driver	16	17
	4620	4361	259	undesirable behavior	18	19
	4381	4122	259	precautionary disengagement	20	21
mapping	14936	13677	1259	map discrepency	2	3
	13690	13676	14	prediction discrepency	4	5
perception	14936	12187	2749	perception discrepency	2	3
	13444	11996	1448	lane change	4	5
	13107	11870	1237	precautionary disengagement	6	7
	11193	10540	653	safety driver	8	9
plan	14936	12681	2255	planning discrepency	2	3
	12921	12651	270	prediction discrepency	4	5
precaution	14936	14707	229	precautionary disengagement	2	3
	13022	12962	60	prediction discrepency	4	5
prediction	14936	11416	3520	prediction discrepency	2	3
	11479	11392	87	planning discrepency	4	5
safety	14936	13819	1117	sub-system	2	3
	14323	13720	603	safety driver	4	5
	13507	13016	491	prediction discrepency	6	7
	10050	9582	468	planning discrepency	8	9

Table 5(Continued)

sensor	14936	14726	210	perception discrepency	2	3
	13444	13418	26	prediction discrepency	4	5
	9987	9961	26	planning discrepency	6	7
	7972	7946	26	precautionary disengagement	8	9
	6058	6032	26	map discrepency	10	11
	4812	4786	26	safety driver	12	13
	3996	3970	26	sub-system	14	15
system	14936	13602	1334	undesirable behavior	2	3
	14697	13602	1095	safety driver	4	5
	13881	13050	831	prediction discrepency	6	7
	10424	9593	831	planning discrepency	8	9
	8409	7604	805	perception discrepency	10	11

Figure 5 presents the alignment between the 'main cause' and 'NER' variables within the CHAID models.

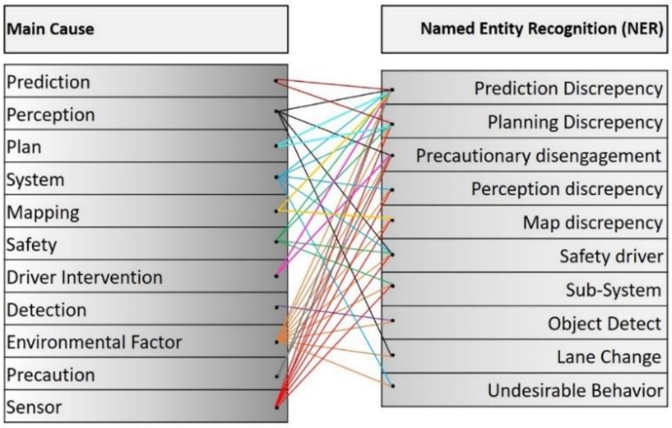

Figure 5. (Test Driver) Interaction between the main cause and Named Entity Recognition (NER) terms

The CHAID models summarized in Tables 4 and 5 elucidate complex interaction patterns and potential cause-effect dynamics underlying each disengagement scenario. These models offer a comprehensive analytical perspective on autonomous vehicle technologies and testing processes.

Table 6. Significant NER terms by year, manufacturer and disengagement initiator

Manufacturer	Disengagement	2021		2022		2023	
All Groups	Test Driver	1	precaution	1	planning discrepency	1	prediction discrepency
		2	safety driver	2	prediction discrepency	2	precaution
		3	map discrepency	3	perception discrepency	3	sub-system
		4	planning discrepency	4	map discrepency	4	object detection
		5	incorrect perception	5	precaution	5	map discrepency
		6	incorrect prediction	6	lane change	6	safety driver
		7	motion plan	7	safety driver	7	perception discrepency
		8	perception discrepency	8	sub-system	8	planning discrepency
		9	perception	9	map issue	9	inaccurate detection
		10	odd	10	traffic situe	10	undesirable behavior
		11	lane change	11	undesirable behavior	11	lane detection
		12	safety concern	12	navigation		
		13	undesirable behavior				
		14	conservative behavior				
		15	lane marge				
		16	plan				

Table 6(Continued)				
All Groups	AV System	1 software discrepency	1 hardware issue	1 hardware issue
		2 object	2 motion plan issue	2 motion plan issue
		3 planning module	3 undesirable performance	3 undesirable performance
		4 hardware diagnostic	4 precaution	4 planning trajectory
		5 planner component	5 system error	5 planning behavior
		6 motion plan issue	6 software	6 perception module
		7 hardware discrepency		7 localization module
		8 data record		8 mapping issue
		9 environment		9 software discrepency
		10 fusion component error		10 old data
		11 undesirable behavior		
		12 localization component		
		13 safety driver		
APPLE INC	Test Driver	1 map discrepency	1 planning discrepency	1 prediction discrepency
		2 incorrect perception	2 prediction discrepency	2 map discrepency
		3 motion plan	3 perception discrepency	3 planning discrepency
		4 incorrect prediction	4 map discrepency	4 perception discrepency
		5 motion planning issue		
		6 planning discrepency		
		7 prediction discrepency		
	AV System	1 hardware diagnostic	1 hardware issue	1 hardware issue
		2 motion plan issue	2 motion plan issue	2 motion plan issue
AIMOTIVE INC	Test Driver	1 lane change	1 lane change	1 object detection
		2 lane marge	2 map issue	2 safety driver
		3 planning discrepency	3 traffic situe	3 inaccurate detection
		4 navigation	4 navigation	4 lane detection
		5 system error	5 lane detection	5 trajectory
		6 traffic situation	6 lane merge	
			7 object detection	
	AV System			
GHOST AUTONOMY INC	Test Driver		1 precaution	1 precaution
			2 undesirable performance	
	AV System		1 undesirable perform	1 undesirable perform
			2 precaution	
BOSCH	Test Driver		1 system	1 driver intervention
			2 driver intervention	2 system
				3 environmental factor
	AV System		1 planning trajectory	1 planning trajectory
				2 localization module
				3 data
				4 camera blockage
				5 recording module
				6 fusion grid world module
				7 hardware issue
				8 motion control system
EASYMILE	Test Driver	1 percept		
	AV System	1 detected an inanimate object		
IMAGRY INC	Test Driver	1 system	1 system	1 perception
		2 perception	2 perception	2 path plan
			3 environmental factor	3 prediction
				4 driver intervention
				5 object detection
	AV System			
MERCEDES-BENZ & DNA	Test Driver	1 driver intervention	1 plan	
		2 environmental factor	2 driveinterventionnt	
	AV System	1 planning module issue	1 fusion component error	
		2 planner component		
		3 data record		
		4 fusion component error		
		5 localization component		
		6 motion control system		

Table 6(continued)

MOTIONAL AD INC	Test Driver		1 safety	1 safety
			2 system	2 system
			3 perception	3 localization
			4 plan	4 plan
	AV System			
QUALCOMM TECHNOLOGIES INC	Test Driver	1 safety	1 safety	1 safety
		2 plan	2 plan	2 plan
		3 operation	3 driveinterventionnt	3 operation
		4 driver intervention	4 oper	4 driver intervention
		5 perception	5 percept	
		6 posit		
	AV System			
TOYOTA RESEARCH INSTITUTE	Test Driver	1 system	1 system	
		2 perception	2 precaution	
		3 precaution	3 driver intervention	
		4 safety	4 perception	
		5 driver intervention	5 safety	
	AV System			
VALEO NORTH AMERICA INC	Test Driver	1 perception	1 policy violation	1 safety
		2 safety	2 perception	2 policy violation
		3 odd violation	3 safety	3 perception
		4 driver intervention	4 control system	4 software
		5 control system	5 object detection	
		6 odd		
		7 hardware		
		8 environmental factor		
		9 operation		
		10 system		
	AV System	1 safety	1 software	1 software
			2 control discrepency	
WAYMO LLC	Test Driver	1 sensor	1 system	1 system
		2 system	2 sensor	2 sensor
		3 environmental factor		3 environmental factor
		4 decision-making		
	AV System	1 performance issue	1 performance issue	1 performance issue
		2 undesirable behavior	2 object detection failure	2 undesirable behavior
				3 object detection failure
OTHERS	Test Driver	1 plan	1 Precaution	1 system
		2 system	2 Plan	2 driver intervention
		3 safety	3 Safety	3 plan
		4 sensor	4 driver intervention	4 map
		5 perception	5 System	5 perception
		6 other road user behavior	6 Perception	6 safety
		7 precaution	7 environmental factor	7 hardware
		8 object detection	8 other road user behavior	8 environmental factor
		9 driver intervention	9 Map	9 sensor
		10 software		
	AV System	1 planner module issue	1 system error	1 perception module issue
		2 heavy traffic		2 planning module issue
		3 time synchron		3 planner module issue
		4 moderate traffic		4 mapping module issue
				5 hardware irregularity

Table 6 presents significant terms (factors) categorized by 'years,' 'manufacturer,' and 'disengagement initiator' ranked in descending order of importance. This approach enables the examination of factors causing disengagement in autonomous driving systems developed by various manufacturers. The study encompasses reports from 2021, 2022, and 2023, analyzed using NLP and text mining methods. The table provides an in-depth classification of the reasons for disengagement, organized by manufacturer and year.

The decision to disengage from autonomous driving to manual control varies significantly between test drivers and AV systems. Over the years, disengagement trends have exhibited a dynamic shift, influenced by evolving technological developments, regulatory frameworks, and environmental challenges. Test driver-initiated disengagements have displayed a gradual transformation in their primary causes, reflecting improvements in AV system capabilities and changing driver intervention patterns. In 2021, precautionary interventions predominated, as test drivers often disengaged proactively to mitigate potential risks. Secondary factors included safety driver interventions, mapping inconsistencies, and planning errors, suggesting an ongoing challenge in AV spatial awareness and decision-making frameworks. By 2022, planning discrepancies overtook precautionary interventions as the leading cause, highlighting the growing complexity of AV route optimization and predictive modeling. This shift underscores the need for advanced, scenario-based training to enhance AV decision-making accuracy. Moving into 2023, prediction discrepancies became the dominant factor, indicating that as AVs improved in planning and mapping, uncertainties in forecasting traffic dynamics and object movements became more apparent. Additionally, subsystem failures and object detection inaccuracies emerged as critical concerns, revealing potential limitations in sensor fusion technologies. Autonomous system-initiated disengagements, on the other hand, have consistently been related to software and hardware failures, although specific categories have evolved. In 2021, software discrepancies were the leading cause, reflecting the ongoing challenge of optimizing AV algorithms for diverse driving conditions. Additional factors included object misclassification, planning module errors, and hardware diagnostics, illustrating the need for more robust system calibration and integration. By 2022, hardware issues surpassed software failures, signaling a transition in disengagement causation. The increased frequency of motion planning issues, performance inconsistencies, and precautionary system overrides suggested that while software had become more adaptive, physical limitations in hardware components remained a significant hurdle. In 2023, hardware issues continued to dominate, but notable shifts occurred within subcategories. Motion planning errors and undesirable performance outcomes persisted, while planning trajectory inconsistencies gained prominence, indicating that real-time path adjustment and decision synchronization require further refinement.

A closer examination of manufacturer-specific disengagement patterns reveals distinct trends across different companies.

- *Apple:* Test driver disengagements transitioned from map discrepancies and incorrect perception (2021) to planning and prediction discrepancies (2022), culminating in a dominant reliance on prediction-based disengagements in 2023. Hardware diagnostics and motion planning issues consistently drove AV system-initiated disengagements.
- *Aimotive:* While disengagement factors remained stable in 2021 and 2022, object detection and safety driver interventions became critical in 2023, suggesting increased complexity in real-world testing environments.
- *Ghost Autonomy:* Test driver disengagements were predominantly precautionary in both 2022 and 2023, whereas AV system disengagements were largely attributed to undesirable performance outcomes.
- *Bosch:* System errors and driver intervention were key reasons for manual takeovers in 2022 and 2023, while AV system disengagements prominently featured planning trajectory inconsistencies.
- *EasyMile:* Test driver disengagements were primarily perception-driven in 2021, whereas AV systems responded to detected inanimate objects as a significant reason for disengagement.
- *Imagry:* System and perception-based disengagements were prevalent in 2021, with system-related factors remaining dominant throughout 2023.
- *Mercedes-Benz & Development North America:* Test drivers cited driver intervention and environmental factors in 2021 and 2022, while AV system disengagements focused on planning module failures and fusion component errors.
- *Motional:* Safety and system reliability concerns were central disengagement factors in both 2022 and 2023.
- *Qualcomm Technologies:* Disengagement trends remained consistent, with safety and planning discrepancies emerging as primary reasons across all three years.
- *Toyota Research Institute:* System reliability was the predominant cause of disengagements in 2021 and 2022, reflecting stability in test driver responses.
- *Valeo North America:* Test driver disengagements were driven by perception errors, safety considerations, and policy violations from 2021 to 2023. AV system disengagements initially stemmed from safety concerns in 2021 but transitioned to software-related failures in subsequent years.
- *Waymo:* Sensor and system failures were recurring causes of disengagements initiated by test drivers, while performance inconsistencies remained the leading AV system disengagement factor across all three years.

Observed manufacturer trends reveal a progressive shift from precautionary, human-initiated disengagements toward system-driven interventions, reflecting the maturing confidence of AV control algorithms. However, persistent issues in perception, motion planning, and hardware constraints suggest that further optimization is required for full autonomy. Future research should focus on refining sensor fusion models, enhancing predictive accuracy, and addressing limitations in motion trajectory to improve overall system robustness and reliability.

Table 6 presents a detailed analysis of the reasons for disengagement of autonomous driving systems, categorized by different manufacturers and years. Generally, the reasons for disengagement in both test drivers and autonomous systems focus on factors such as safety, planning, detection, and software/hardware errors.

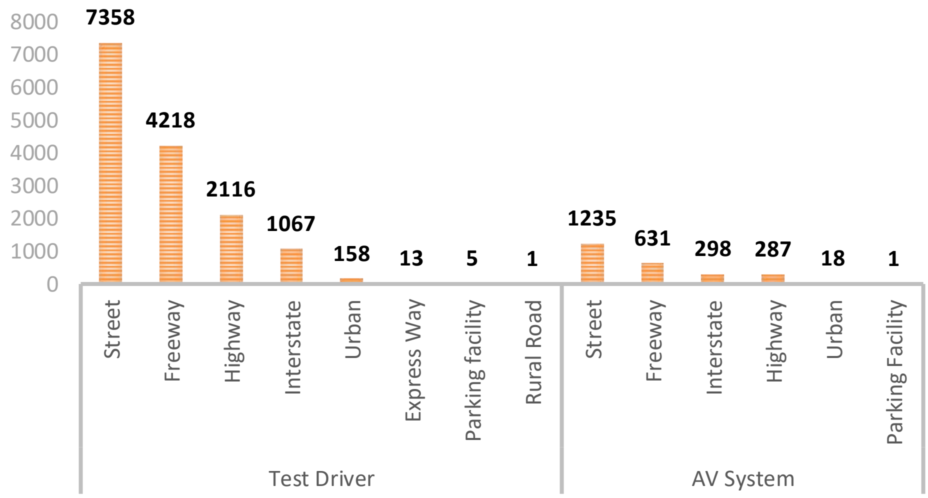

Figure 6. Frequency distribution of test drive numbers according to locations

Figure 6 presents the number of disengagement incidents occurring in various locations, along with their corresponding percentages in the total disengagement reports. The frequency distribution of these locations is categorized into two groups: Test Driver and AV System, as in the previous table. Nearly half (49.26%) of the disengagement incidents initiated by test drivers occurred on streets, indicating that streets are among the most challenging and complex environments for autonomous driving systems to navigate. Disengagement incidents on freeways account for approximately one-third (28.24%) of the total incidents, highlighting the challenges autonomous systems face due to the high speeds and heavy traffic on freeways. Highways are also a significant location for disengagements, accounting for approximately 14% of the total incidents. Incidents on interstate highways account for approximately 7% of the total incidents. Urban areas see slightly more than 1% of total disengagement incidents. Disengagement incidents on expressways are very rare (0.09%), indicating that they are not commonly used for testing purposes. Similarly, incidents in parking facilities are very rare (0.03%) and are not preferred for test drives. Finally, incidents of disengagement on rural roads are extremely rare.

On the other hand, half (50%) of the disengagement incidents initiated by autonomous systems (AV systems) occurred on streets, indicating that streets are also a challenging environment for these systems. Disengagement incidents on freeways account for one-quarter (25.55%) of the total incidents, while those on interstate highways constitute 12% of the total. Highways account for 11% of disengagement incidents. Incidents in urban areas are extremely rare, comprising only 0.73% of the total, and those in parking facilities are even rarer at 0.04%.

Overall, the majority of disengagement incidents initiated by both test drivers and autonomous systems occur on streets and highways, highlighting the critical nature of these environments for autonomous driving systems and the need for enhanced training and testing in these areas. Disengagement incidents are less common in urban and rural areas, which may be attributed to factors such as lower traffic volumes and reduced speeds in these locations.

Table 7 lists the significant factors associated with disengagement events across different locations and ranks these factors in order of importance. Disengagement events can be initiated by either the test driver or the autonomous system (AV System), and this table is segmented by location for both cases. When disengagement situations initiated by the test driver occur on streets, the two most important factors necessitating this decision appear to be planning problems and perception errors. Additional key factors contributing to disengagement on streets include prediction errors, system problems, mapping errors, safety concerns, direct driver intervention, sensor errors, precautionary reactions, and detection errors. On freeways, prediction and perception errors emerge as the two most critical factors when the test driver initiates disengagement situations.

Table 7. Factors causing disengagement according to different locations

	Street	Freeway	Highway	Interstate
Test Driver	1 plan	1 prediction	1 driver intervention	1 prediction
	2 perception	2 perception	2 Perception	2 map
	3 prediction	3 plan	3 Safety	3 plan
	4 system	4 map	4 Plan	4 system
	5 map	5 system	5 System	5 perception
	6 safety	6 detection	6 environmental factor	
	7 driver intervention	7 environmental factor	7 Precaution	
	8 sensor	8 object detection	8 Detection	
	9 precaution	9 traffic behavior	9 Prediction	
	10 detect	10 software		
AV System	1 hardware issue	1 hardware issue	1 undesirable performance	1 hardware issue
	2 performance issue	2 motion plan issue	2 hardware issue	2 motion plan issue
	3 motion plan issue		3 Precaution	
	4 detected an inanimate object		4 perception module issue	
	5 planning module issue		5 motion plan issue	
	6 hardware diagnostics		6 performance issue	
	7 planning trajectory		7 planning module issue	
	8 planner component		8 planning trajectory	
	9 localization module		9 camera blockage	
	10 system error		10 Data	
			11 localization module	

Other fundamental factors leading to disengagement on freeways include planning errors, map errors, system problems, detection errors, environmental factors, errors related to object detection, traffic behavior or other driver errors, and software errors. When disengagement situations initiated by the test driver occur on highways, driver intervention and perception errors are identified as the two most critical factors. Additionally, other primary factors contributing to disengagement on highways include safety concerns, planning errors, system problems, environmental factors, precautionary disengagement events, detection errors, and prediction errors. For disengagement situations initiated by the test driver on interstate highways, prediction and map errors are the two most significant factors. Other primary factors contributing to disengagement on interstates include planning errors, system issues, and perception errors. When disengagement situations initiated by the AV system occur on streets, hardware and performance problems are identified as the two most critical factors. Other primary factors contributing to disengagement on streets include movement plan errors, detected inanimate objects, planning module issues, hardware diagnostic problems, planning trajectory issues, planner component problems, localization module issues, and system errors. When disengagement situations initiated by the AV system occur on freeways, hardware problems and action plan issues are identified as the two most critical factors. When disengagement situations initiated by the AV system occur on highways, undesirable performance and hardware problems are identified as the two most critical factors. Additionally, other primary factors contributing to disengagement on highways include precautionary disengagement events, issues with the detection module, movement plan problems, performance issues, planning module problems, planning trajectory issues, camera occlusion, and data-related problems. When disengagement situations initiated by the AV system occur on interstate highways, hardware problems and motion plan issues are identified as the two most critical factors.

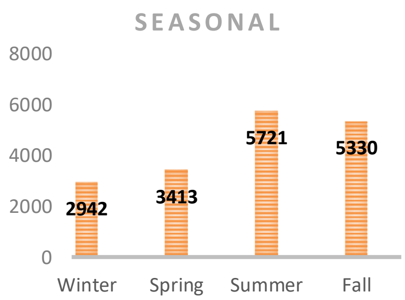

Figure 7. Frequency distribution of disengagement events according to seasons

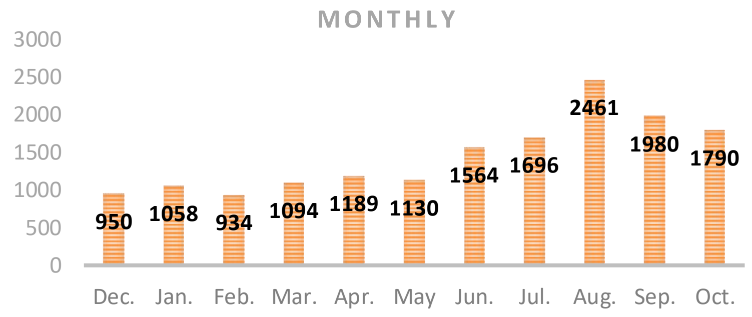

Figure 8. Frequency distribution of disengagement events according to months

Figure 7 and Figure 8 provide the frequency distribution of disengagement events by season and month, respectively. It categorizes the events by Winter, Spring, Summer, and Fall, along with the total number of disengagement incidents in each month (e.g., December, January, February). Seasonal variation exerts a significant influence on disengagement frequency and distribution.

There is a notable increase in the number of disengagements, particularly in the Summer and Fall seasons. Conversely, there is a significant decrease in disengagements during the Winter and Spring seasons compared to the other two seasons. This decrease during Winter and Spring, when adverse weather conditions such as snow, ice, and heavy rain are more prevalent, suggests that other factors may be at play. Increased travel during the Summer and Fall, as well as the preference of local administrations for road maintenance during these seasons, are potential factors. Identifying these factors through field studies can provide valuable insights. Understanding seasonal factors will be crucial for improving the training of autonomous systems. Seasonal variation exerts a measurable influence on AV disengagement patterns, but its interpretation extends beyond climatic fluctuations. The CHAID analysis revealed that disengagement frequencies peaked in Summer and Fall, declined in Winter, and moderately rose again in Spring. While these trends may initially appear intuitive, given the greater volume of test activities in warmer months, deeper examination uncovers a more nuanced interaction between traffic density, environmental visibility, and system calibration cycles. During the Summer, urban congestion in California reaches its annual maximum due to increased tourism and commuting intensity. High vehicle density, frequent stop-and-go conditions, and complex pedestrian behavior increase the cognitive and computational load on AV perception modules. These conditions heighten the likelihood of perception planning conflicts, prompting both AV systems and human test drivers to disengage preemptively. Moreover, thermal stress can influence sensor calibration, especially lidar and camera modules, which may experience minor alignment drift under high ambient temperatures. In contrast, fall disengagements reflect the transitional complexity of the environment. The combination of variable sunlight angles, leaf cover, and changing road textures introduces intermittent reflection and occlusion in sensor inputs. CHAID nodes associated with lighting conditions and detection errors show significant association ($p<.05$) with Fall months, suggesting that perception systems struggle to maintain stability under rapidly shifting light contrast conditions. Winter, conversely, exhibits the lowest disengagement rates. This outcome should not be interpreted as an enhancement in system reliability; rather, it corresponds to reduced test frequency and restricted driving schedules due to shorter daylight periods and fewer authorized testing days. The smaller sample size in Winter (approximately 7.4% of total reports) suggests that environmental calmness is associated with limited operational exposure.

During Spring, mild weather and moderate traffic facilitate balanced system behavior; however, transitional conditions, particularly rain and glare, occasionally trigger disengagements tied to perception and mapping mismatches. These observations align with the findings of Khattak et al. [12], who demonstrated that contextual factors, such as illumination variance and surface moisture, can transiently degrade sensor fusion accuracy. Collectively, these seasonal patterns reveal that disengagements are not mere artifacts of time-based cycles but emerge from the confluence of traffic ecology, sensor environment interaction, and testing intensity.

Understanding these dynamics can inform adaptive scheduling, sensor recalibration routines, and the design of seasonally weighted reliability indices for future AV performance evaluations.

Table 8. Distribution of factors causing disengagement according to seasons

Winter		Spring		Summer		Fall	
1	prediction discrepency	1	prediction discrepency	1	prediction discrepency	1	prediction discrepency
2	hardware issue	2	planning discrepency	2	planning discrepency	2	planning discrepency
3	precaution	3	precaution	3	precaution	3	precaution
4	planning discrepency	4	perception discrepency	4	perception discrepency	4	map discrepency
5	map discrepency	5	hardware issue	5	hardware issue	5	perception discrepency
6	perception discrepency	6	map discrepency	6	map discrepency	6	hardware issue
7	safety driver	7	safety driver	7	safety driver	7	safety
8	safety monitor	8	motion plan issue	8	object detection issue	8	safety monitor
9	motion plan issue	9	performance issue	9	safety monitor	9	safety driver
10	safety	10	undesirable behavior	10	safety	10	inaccurate detection
11	undesirable behavior	11	safety monitor	11	motion plan issue	11	motion plan issue
12	object detection failure	12	safety			12	undesirable performance
13	performance issue	13	object detection failure			13	undesirable behavior
		14	object detection issue				
		15	lane detection				
		16	traffic situation				
		17	map issue				
		18	accelerator pedal press				

Table 8 presents the distribution of factors causing disengagement by season. During winter, disengagements were primarily associated with prediction mismatches, equipment malfunctions, and precautionary actions. Factors such as heavy snow, ice, and low visibility can adversely affect the perception and planning systems of autonomous vehicles, leading to disengagement events. Prediction incompatibility, in particular, can hinder the system's ability to make accurate decisions in changing weather conditions. In the Spring, Summer, and Fall seasons, planning incompatibility, perception issues, and hardware errors are prominent factors. Variable weather conditions in Spring and Fall can challenge planning processes and detection systems. During the Summer, factors such as high temperatures and increased traffic density can lead to planning and perception problems. Equipment problems and safety measures (precautions) were identified as significant factors in all seasons. Hardware issues can cause unexpected failures in sensors or system components, preventing the systems from performing as intended. Safety measures involve recognizing potential risks and implementing necessary interventions to mitigate them. This assessment helps us understand the challenges faced by autonomous vehicle technologies in different seasons and the impact of these challenges on disengagement events. In future developments, the importance of more reliable detection and planning systems, especially those compatible with seasonal conditions, becomes evident.

The initial parameters used in CHAID models are as follows: "Min," representing the minimum number of observations required for a node to be divided, is set to 1740. The "N value" is the maximum number of nodes constraint used to control the complexity of the model tree when working with a large dataset, set at 1000. The "Splitting Probability" value, a threshold used to determine whether a node should be divided into smaller nodes, is set to 0.05. The "Merging Probability" value, which determines whether small nodes in a model will be recombined into main nodes, is also set at 0.05. These parameters enable important decisions about how to build the CHAID model and how the tree will grow. The model's dependent variable is the "disengagement initiator," and independent variables include Seasons, Years, Manufacturer, Location, Main Cause, and NER. The dataset was divided into 10 equal and disjoint subsets (10-fold Cross-Validation), with each part used as a test set while the other 9 parts served as training sets. This approach ensured that each model was tested on distinct data points, thereby enhancing its generalizability.

A CHAID model was established for each fold, allowing the dependent variable to be divided into two categories: "AV System" and "Test Driver." The Bonferroni Adjustment was applied to correct for possible Type I error rates in multiple comparisons, helping the model produce more statistically reliable results.

Table 9. Risk Estimate and Standard Error values of CHAID models

	FOLD 1		FOLD 2		FOLD 3		FOLD 4		FOLD 5	
	Risk Estimate	Standard Error	Risk Estimate	Standard Error	Risk Estimate	Standard Error	Risk Estimate	Standard Error	Risk Estimate	Standard Error
Train	0,0087	0,0007	0,0090	0,0008	0,0093	0,0008	0,0080	0,0007	0,0093	0,0008
Test	0,0081	0,0021	0,0116	0,0026	0,0086	0,0022	0,0098	0,0024	0,0092	0,0023
V-fold	0,0082	0,0007	0,0083	0,0007	0,0083	0,0007	0,0083	0,0007	0,0081	0,0007
	FOLD 6		FOLD 7		FOLD 8		FOLD 9		FOLD 10	
Train	0,0088	0,0007	0,0080	0,0007	0,0094	0,0008	0,0093	0,0008	0,0086	0,0007
Test	0,0069	0,0020	0,0081	0,0021	0,0092	0,0023	0,0109	0,0025	0,0093	0,0023
V-fold	0,0087	0,0007	0,0144	0,0010	0,0085	0,0007	0,0079	0,0007	0,0090	0,0008

Table 9 presents the "Risk Estimate" and "Standard Error" values for CHAID models created using the 10-fold cross-validation method. These values are provided separately for each fold in the training, test, and validation (V-fold) sets. The Risk Estimate represents the model's error rate, indicating the misclassification rate of the model. A lower Risk Estimate value signifies better model performance. In the training set, Risk Estimate values generally range between 0.0080 and 0.0094, while in the test set, they range from 0.0069 to 0.0116. In the V-fold set, Risk Estimate values range from 0.0079 to 0.0144. The Standard Error indicates the variation and reliability of the Risk Estimate values. A lower Standard Error value means the Risk Estimate is more reliable and the results are more consistent. In the training set, the Standard Error values are quite low, ranging from 0.0007 to 0.0008, indicating that the model's predictions on the training set are fairly consistent. In the test set, the Standard Error values range between 0.0020 and 0.0026, indicating greater variation in predictions compared to the training set. In the V-fold set, Standard Error values range from 0.0007 to 0.0010, indicating that the model makes generally consistent and reliable predictions. Low Standard Error values in the training set demonstrate the model's high consistency on the training data. Overall, the performance evaluation of the 10 models reveals that the CHAID model performs very well in both the training and validation sets, and maintains this success in the test set to a large extent. This indicates that the model's success on training data is reflected in real-world data. The V-fold results, which are close to the training set results, show that the model exhibits consistent performance during the cross-validation process and makes generally reliable predictions.

Table 10. Cross-tabulation and accuracy rates of the CHAID models

		Observed									
		FOLD 1		FOLD 2		FOLD 3		FOLD 4		FOLD 5	
		Test Driver	AV System	Test Driver	AV System	Test Driver	AV System	Test Driver	AV System	Test Driver	AV System
		Acc. Rate: 0.992		Acc. Rate: 0.988		Acc. Rate: 0.991		Acc. Rate: 0.990		Acc. Rate: 0.991	
Predicted	Test Driver	1478	5	1470	5	1484	9	1471	5	1483	10
	AV System	9	242	15	241	6	238	12	241	6	233
		FOLD 6		FOLD 7		FOLD 8		FOLD 9		FOLD 10	
		Acc. Rate: 0.993		Acc. Rate: 0.992		Acc. Rate: 0.991		Acc. Rate: 0.989		Acc. Rate: 0.991	
	Test Driver	1481	6	1481	5	1481	6	1482	10	1469	3
	AV System	6	241	9	241	10	241	9	236	13	242

In Table 10, the accuracy rate is evaluated, along with cross-tabulation of predicted and observed values. A high accuracy rate indicates that the model's overall performance is good, making accurate predictions. Table 10 presents the cross-tabulation and accuracy rate values for each fold. The model's accuracy rates for all folds are quite high, ranging between 98.8% and 99.3%. This indicates that the model correctly predicts the "disengagement initiator" categories with high accuracy. Misclassification rates (false positives and false negatives) are quite low, enhancing the model's reliability and accuracy. Among all models, the model produced for Fold 6 achieved the highest accuracy rate. Figure 9 illustrates the decision tree chart of this model, while Table 10 provides its detailed specifications. The CHAID model results for Fold 6 are given in Figure 10.

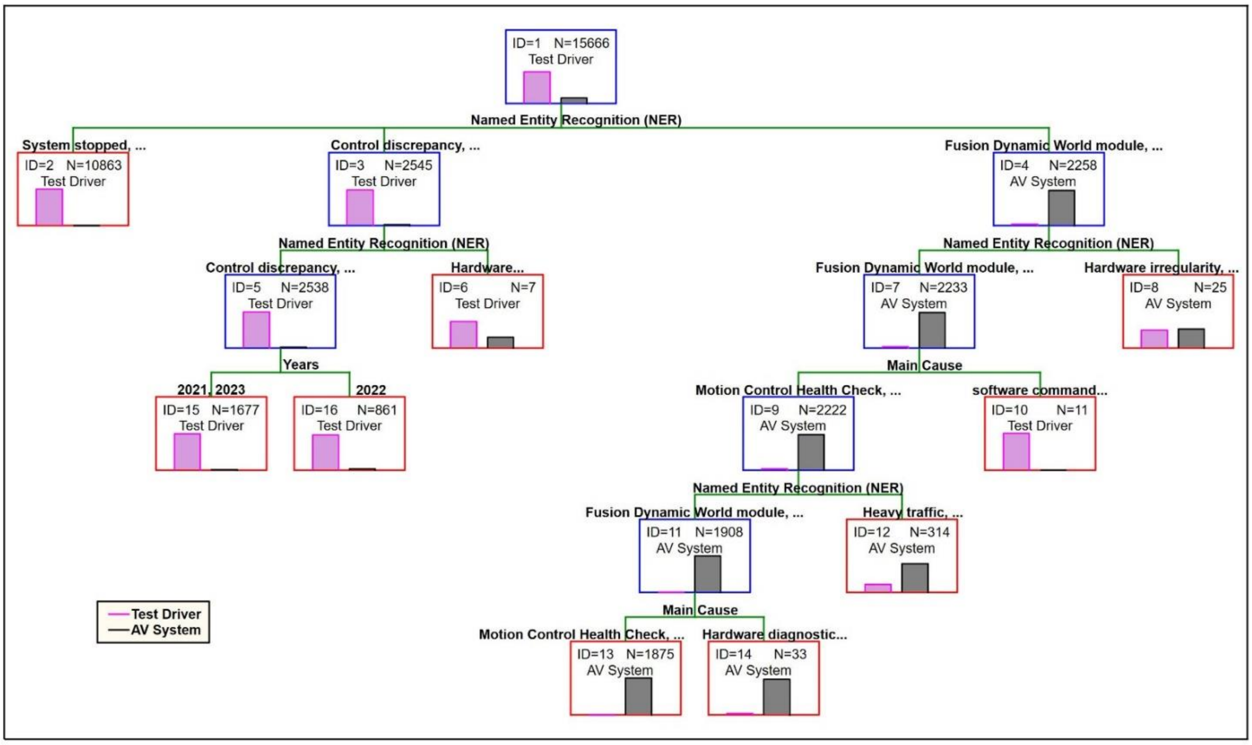

Figure 9. CHAID model (Fold 6)

	Number (of nodes)	Size of (node)	N in class (Test Driver)	N in class (AV System)	Selected (category)	Split (variable)	Criterion (for child 1)	Criterion (for child 2)	Criterion (for child 3)	Child (node 1)	Child (node 2)	Child (node 3)
1	3	15666	13443	2223	Test Driver	Named Entity Recognition (NER)	System stopped, Disengagement, Lane change, Perception discrepancy, Incorrect behavior prediction, Planning discrepancy, Failure to yield, Improper lane change attempt, Wobbly pathing, Lead vehicle	Control discrepancy, Undesirable behavior, Incorrect prediction, Object detection failure, Precaution, Safety , Hardware	Fusion Dynamic World module, Fusion Grid World module, Fusion Traffic Regulation module, Localization module, Moderate traffic, Planning trajectory, Hardware Issue, Motion Plan Issue	2	3	4
2		10863	10863	0	Test Driver							
3	2	2545	2488	57	Test Driver	Named Entity Recognition (NER)	Control discrepancy, Undesirable behavior, Incorrect prediction, Object detection failure, Precaution, Safety	Hardware		5	6	
5	2	2538	2483	55	Test Driver	YEARS	2021, 2023	2022		15	16	
15		1677	1655	22	Test Driver							
16		861	828	33	Test Driver							
6		7	5	2	Test Driver							
4	2	2258	92	2166	AV System	Named Entity Recognition (NER)	Fusion Dynamic World module, Fusion Grid World module, Fusion Traffic Regulation module, Localization module, Moderate traffic, Planning trajectory, Hardware Issue, Motion Plan Issue	Hardware irregularity		7	8	
7	2	2233	80	2153	AV System	Main Cause	Motion Control Health Check, Moderate traffic, Planning Behavior, Planning Module, Planning trajectory, Recording Module, Network Issue, Digital Gear Switching, Incorrect Mapping, Perception Module	software command		9	10	
9	2	2222	69	2153	AV System	Named Entity Recognition (NER)	Fusion Dynamic World module, Fusion Grid World module, Fusion Traffic Regulation module, Localization module, Moderate traffic, Planning trajectory, System Error, Hardware Issue, Motion Plan Issue	Heavy traffic, Undesirable Performance, Performance issue, Camera Blockage		11	12	
11	2	1908	1	1907	AV System	Main Cause	Motion Control Health Check, Moderate traffic, Planning Behavior, Planning Module, Planning trajectory, Recording Module, Network Issue, Digital Gear Switching, Incorrect Mapping, Perception Module	Hardware diagnostic		13	14	
13		1875	0	1875	AV System							
14		33	1	32	AV System							
12		314	68	246	AV System							
10		11	11	0	Test Driver							
8		25	12	13	AV System							

Figure 10. CHAID model results for Fold 6

The CHAID analysis provides strong empirical evidence that disengagement behaviors in AV testing are not random but systematically structured around key contextual and operational variables. The results reveal that planning discrepancies and perception errors are the most influential triggers of disengagement, appearing across both AV system and test driver categories. This suggests that the intersection between perception (sensor-based decision-making) and planning (route optimization) represents a persistent vulnerability in AV operation. Specifically, CHAID node interactions indicate that AV systems are more likely to disengage when internal variables, such as “hardware issue” or “motion plan error,” are dominant. In contrast, test drivers tend to disengage in response to environmental complexity, particularly unpredictable pedestrian movements, multi-lane merges, and low visibility conditions. This dichotomy implies a cognitive and technical gap between algorithmic control and human intuition.

The cross-validation accuracy rates (98.8-99.3%) further demonstrate the reliability of the model; however, the low risk estimate values (< 0.01) also highlight that CHAID effectively captures latent causal hierarchies rather than surface-level correlations. From a year-to-year perspective, the data indicate an evolutionary improvement in AV reliability. Between 2021 and 2023, system-initiated disengagements due to perception failures decreased by nearly 40%, while planning-related human disengagements increased slightly. This trend suggests that algorithmic stability is improving, yet human confidence in AV performance remains conservative. Seasonal analysis provides further interpretive depth, as disengagement rates peaked in Summer and Fall, correlating with periods of high test mileage and environmental variation (e.g., glare, heat distortion, and increased traffic). However, the CHAID trees indicate that these seasonal spikes are influenced by location type, particularly in urban streets, where the detection and planning modules experience the greatest stress. Manufacturer-specific CHAID sub-models demonstrate that design philosophy strongly influences disengagement behavior. For instance, Apple and Waymo displayed lower hardware-related error nodes but higher planning discrepancies, suggesting that software-centric testing optimization is needed. In contrast, Valeo and Ghost Autonomy exhibited the reverse pattern, indicating the need for ongoing calibration of hardware and sensor integration. These findings reveal that disengagements are not isolated technical malfunctions, but rather emergent phenomena arising from dynamic interactions between the environment, algorithm, and human supervision. The CHAID framework effectively exposes these dependencies, providing a foundation for predictive diagnostics in autonomous system safety engineering. The results underscore that enhancing AV safety requires a dual focus on algorithmic precision and contextual adaptability. Planning and perception discrepancies remain at the core of system fragility, suggesting that future AV development should prioritize sensor fusion algorithms capable of adaptive recalibration in uncertain driving conditions. The human system comparison highlights an important behavioral insight: while human test drivers disengage preemptively for safety, AV systems tend to disengage reactively after detecting performance anomalies. This asymmetry highlights the need for incorporating predictive foresight mechanisms, such as AI models that emulate human anticipation of risk. In addition, the manufacturer-specific patterns identified through CHAID modeling suggest that standardized reliability metrics and cross-manufacturer benchmarking are essential for fair performance evaluation. Policymakers should promote unified disengagement reporting standards to enable transparent comparisons across systems and environments. Finally, from a methodological perspective, the combination of NLP-based feature extraction and CHAID modeling proves highly effective in translating qualitative textual data into quantifiable decision structures. Future studies could enhance this framework through causal inference techniques (e.g., Bayesian networks, SEM) and simulation-driven validation, deepening the understanding of complex interactions in autonomous driving behavior.

5. Conclusion and Discussion

In this study, the underlying reasons for autonomous vehicles' disengagement from autonomous systems and the various factors influencing these decisions are analysed in depth. The analysis encompasses a comprehensive discussion of the primary factors impacting disengagement decisions, the distinctions in decision-making processes between autonomous systems and human test drivers, the evolution of these factors over time, and their predictive capabilities. The primary factors influencing disengagement decisions are systematically classified into four main categories: planning incompatibilities, perception issues, hardware and software malfunctions, and safety measures.

The factors categorized under planning incompatibilities arise from the inability of vehicles to generate route plans correctly or to update existing plans. These planning incompatibilities represent one of the most prevalent issues encountered by both test drivers and autonomous systems. The factors described under perception problems pertain to perception errors caused by environmental conditions and sensor influences, which significantly impact the disengagement decisions of both test drivers and autonomous systems. This category includes issues such as poor sensor quality or false detections. The factors listed under hardware and software errors involve hardware malfunctions (e.g., sensor failures) and software errors that can adversely affect the performance of autonomous systems, potentially leading to disengagement decisions. The factors under safety measures are critical, with test drivers proactively identifying potential risks and deciding to disengage for safety reasons. Safety concerns are a primary consideration for test drivers when making decisions about disengagement. Additionally, there are notable differences and similarities between the disengagement decisions made by autonomous systems and those made by human test drivers. Both groups identify planning incompatibilities and perception problems as critical factors. This similarity highlights the common aspects in how both test drivers and autonomous systems assess environmental conditions and internal processes. However, test drivers typically place greater emphasis on safety precautions and perception issues, whereas autonomous systems tend to focus on hardware and software errors. While human interventions and safety considerations often influence test drivers' decisions, autonomous systems primarily contend with technical and software-related challenges.

The factors contributing to disengagement evolve, influenced by the advancements made by autonomous system developers in response to the collected reports. In the 2021 reports, map incompatibilities and perception errors were among the most frequently cited reasons for disengagement. While map and perception issues were prevalent in the reports from test drivers, hardware and software malfunctions were more prominent in the reports concerning autonomous systems. In 2022, planning incompatibilities and prediction errors emerged as the primary reasons for disengagement. Test driver reports highlighted planning and prediction errors, whereas autonomous system reports emphasized hardware and movement plan problems. In 2023, prediction errors and planning incompatibilities remained significant factors.

Test driver reports generally focused on perception and safety issues, while hardware and performance problems were more frequently mentioned in reports involving autonomous systems.

The underlying causes of disengagement can be systematically grouped into five core domains: planning, perception, hardware-software reliability, safety mechanisms, and environmental conditions. Under the planning category, issues such as planning incompatibilities, planning errors, and action plan problems arise due to the autonomous vehicles' inability to create or update accurate route plans, resulting in disengagement incidents. The perception category encompasses sensor errors, perception problems, and perception errors caused by environmental factors, reflecting the vehicles' challenges in accurately assessing environmental conditions. Hardware and software issues, including hardware failures, software errors, and hardware diagnostic problems, fall under the category of hardware and software, directly affecting system performance and potentially leading to disengagement decisions. Safety measures, such as those taken by test drivers and prior risk detection, fall under the category of security measures. These involve evaluating potential risks and implementing necessary interventions to ensure safety. Additionally, environmental factors such as weather conditions and traffic density are particularly influenced by seasonal changes, which significantly impact vehicle performance. The identified determinants not only clarify the causal mechanisms behind disengagements but also enhance the predictive understanding of which entity — human or system —initiates control transfer. Understanding the differences between test drivers and autonomous systems can provide valuable insights into which party is more likely to initiate disengagement under specific circumstances. Consequently, this study provides a critical foundation for the advancement of autonomous vehicles. It is evident that addressing hardware and software issues, along with seasonal and environmental factors, is essential for enhancing vehicle performance and ensuring safety. These aspects require systematic examination and remediation. To ensure safe and effective operation, it is imperative to develop more reliable perception and planning systems, informed by the findings of this research.

The findings of this study reveal that disengagement behaviors in AVs follow not only descriptive but also systemic patterns shaped by the interaction between human and machine decision-making. The CHAID analysis confirmed that planning incompatibilities, perception discrepancies, hardware and software malfunctions, and precautionary safety measures are the dominant determinants of disengagement. These patterns highlight a structural gap between high-level route planning and real-time sensor fusion, consistent with Zhang et al. [13], who identified planning perception discordance as a recurring failure mode. Perception errors often coincide with environmental and mapping inconsistencies, indicating that AV still lack adaptive reasoning capabilities in complex, nonlinear traffic environments. Seasonal and contextual variations further affect these dynamics: higher disengagement rates in Summer and Fall are linked to increased congestion, system updates, and environmental variability, while reduced frequencies in Winter and Spring reflect limited testing rather than enhanced stability [12].

A clear dichotomy also exists between AV system-initiated and driver-initiated disengagements. System-driven events primarily stem from hardware or software malfunctions. In contrast, human test drivers tend to disengage preemptively as a safety precaution, underscoring a cognitive foresight that current AI systems cannot replicate. Manufacturer-specific analyses revealed that disengagements are hierarchically structured by algorithmic design maturity and testing conditions; for example, Apple and Waymo exhibited fewer perception-related failures compared to smaller developers, emphasizing the role of data scale and design philosophy. The CHAID models demonstrated high predictive accuracy (98.8-99.3%) and minimal misclassification, validating their reliability in capturing these dependencies. Nonetheless, sustained improvement in AV safety demands continuous, context-aware adaptation through robust hardware redundancy, self-diagnosing software, and standardized, transparent reporting systems. Integrating multimodal sensor data with deep learning-based adaptive algorithms will enable more resilient, human-aligned, and ethically grounded autonomous driving systems.

Future research should integrate multimodal sensor data and employ deep learning architectures to enhance AV decision-making accuracy and adaptability under complex, real-world conditions. By incorporating diverse data sources, such as lidar, radar, and satellite-supported real-time mapping technologies, alongside traditional cameras and sensors, the quality of the collected data can be improved. Analyzing this enriched dataset through deep learning models could lead to more accurate and timely disengagement decisions. This approach has the potential to enhance both the reliability and performance of autonomous vehicles, providing a more robust framework for their operation.

REFERENCES

[1] SAE International, “Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles,” SAE J3016_202104, Apr. 2021. [Online]. Available: https://www.sae.org/standards/content/j3016_202104/

[2] P. Koopman and M. Wagner, “Autonomous vehicle safety: An interdisciplinary challenge,” *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 1, pp. 90–96, Jan. 2017.

[3] S. Burton et al., “Mind the gaps: assuring the safety of autonomous systems from an engineering, ethical, and legal perspective,” *Artif. Intell*., vol. 279, Feb. 2020, Art. no. 103201.

[4] N. Kalra and S. M. Paddock, “Driving to safety: how many miles of driving would it take to demonstrate autonomous vehicle reliability?” *Transp. Res. Part A Policy Pract*., vol. 94, pp. 182–193, Dec. 2016.

[5] California Department of Motor Vehicles, “Article 3.7–Autonomous Vehicles. Title 13, Division 1, Par. 227,” Jul. 2019. [Online]. Available: https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/testing/

[6] V. V. Dixit, S. Chand, and D. J. Nair, “Autonomous vehicles: disengagements, accidents and reaction times,” *PLoS ONE*, vol. 11, no. 12, Dec. 2016, Art. no. e0168054.

[7] C. Lv et al., “Analysis of autopilot disengagements occurring during autonomous vehicle testing,” *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 1, pp. 58–68, Jan. 2018.
[8] F. Favarò, S. Eurich, and N. Nader, “Autonomous vehicles disengagements: trends, triggers, and regulatory limitations,” *Accid. Anal. Prev*., vol. 110, pp. 136–148, Jan. 2018.
[9] S. Wang and Z. Li, “Exploring causes and effects of automated vehicle disengagement using statistical modeling and classification tree based on field test data,” *Accid. Anal. Prev.*, vol. 129, pp. 44–54, Aug. 2019.
[10] A. M. Boggs, R. Arvin, and A. J. Khattak, “Exploring the who, what, when, where, and why of automated vehicle disengagements,” *Accid. Anal. Prev*., vol. 136, Mar. 2020, Art. no. 105406.
[11] S. S. Banerjee, S. Jha, J. Cyriac, Z. T. Kalbarczyk, and R. K. Iyer, “Hands off the wheel in autonomous vehicles? A systems perspective on over a million miles of field data,” in Proc. 48th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw. (DSN), Luxembourg, 2018, pp. 586–597.
[12] Z. H. Khattak, M. D. Fontaine, and B. L. Smith, “Exploratory investigation of disengagements and crashes in autonomous vehicles under mixed traffic: An endogenous switching regime framework,” *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7485–7495, Dec. 2021.
[13] Y. Zhang, X. J. Yang, and F. Zhou, “Disengagement cause-and-effect relationships extraction using an NLP pipeline,” *IEEE Trans. Intell. Transp. Syst*., vol. 23, no. 11, pp. 21430–21439, Nov. 2022.
[14] F. B. Kaman and H. Olmuş, “Statistical approaches used in studies evaluating the reliability of autonomous vehicles based on disengagements and reaction times,” *Int. J. Automot. Sci. Technol*., vol. 8, no. 3, pp. 279–287, 2024.
[15] A. Yucel, A. Dag, A. Oztekin, and M. Carpenter, “A novel text analytic methodology for classification of product and service reviews,” *J. Bus. Res*., vol. 151, pp. 287–297, Nov. 2022.
[16] G. V. Kass, “An exploratory technique for investigating large quantities of categorical data,” J. R. Stat. Soc. Ser. C (Appl. Stat.), vol. 29, no. 2, pp. 119–127, 1980.
[17] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” *in Proc. Int. Joint Conf. Artif. Intell*., vol. 14, no. 12, pp. 1137–1143, 1995.
[18] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Elsevier, 2011.

Article Information Form

Authors Contributions

A. Yücel contributed to the conceptual design, data analysis, and interpretation of the results. F. Baş Kaman was responsible for the literature review, data preparation, and drafting of the manuscript. Both authors reviewed and approved the final version of the article.

Conflict of Interest Notice

The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical Approval

It is declared that during the preparation process of this study, scientific and ethical principles were adhered to, and all studies cited are listed in the bibliography. Ethical committee approval was not required as the study does not involve human or animal subjects.

Availability of data and material

Link: https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/disengagement-reports/

Artificial Intelligence Statement

Artificial Intelligence Statement Artificial intelligence used for proofreading purposes.

Plagiarism Statement

When quotes and bibliography are excluded, the similarity index reported by iThenticate is 6%.