INTERNATIONAL JOURNAL OF AUTOMOTIVE SCIENCE AND TECHNOLOGY

2024, VOL. 8, NO:3, 279-287 www.ijastech.org e-ISSN: 2587-0963

Statistical Approaches Used in Studies Evaluating the Reliability of Autonomous Vehicles Based on Disengagements and Reaction Times

Ferhan Baş Kaman[1*] Hülya Olmuş[2]

[1.] Department of Finance and Banking, Şereflikoçhisar Faculty of Applied Science, Ankara Yıldırım Beyazıt University, Ankara, Turkey
[2.] Department of Statistics, Faculty of Science, Gazi University, Ankara, Turkey.

Review Article

History
Received 07.05.2024
Revised 24.07.2024
Accepted 15.08.2024

Contact
* Corresponding author
Ferhan Baş Kaman
ferhanbaskaman@aybu.edu.tr
Address: Department of Finance and Banking, Şereflikoçhisar Faculty of Applied Science, Yıldırım Beyazıt University, Ankara, Turkiye
Tel:+903129062531

Abstract

The rapid development of autonomous vehicle (AV) technology highlights the critical importance of enhancing the reliability of these vehicles. Due to the need to test the reliability of AVs, since 2014, the California Department of Motor Vehicles has permitted autonomous vehicle manufacturers to establish an AV Testing program, enabling them to test automated systems on the transportation network. With this, studies on the reliability of AVs have increased rapidly. The most emphasized issues regarding the reliability of AVs have been disengagements, accidents, and reaction times. In this study, disengagements and reaction times are categorized and explained in detail according to the data type, company, period, and statistical method. The data used in the studies cover the years 2014-2020. When examining studies on the reliability of AVs, until 2018, inferences were generally made using real data and descriptive statistics, particularly with methods such as correlation analysis and calculation of disengagements per mile, which investigates the relationship between distance traveled and disengagements. However, since 2018, machine learning has gained importance in evaluating AV reliability. It has been observed that regression, classification, and decision trees were frequently used during this period. Techniques such as deep transfer learning, text mining, and natural language processing also stand out. Furthermore, Software Reliability Growth Models are used to measure software reliability, playing an essential role in evaluating, analyzing, and improving the performance and reliability of AVs. This study aims to reveal the development and diversity of the statistical methods used to determine AV reliability. Additionally, this study aims to guide and provide insights to researchers in the field about the statistical approaches they can utilize.

Keywords: Autonomous vehicle reliability; Disengagements; Reaction times; Statistical methods

To cite this paper: Baş Kaman, F., Olmuş, H. Statistical Approaches Used In Studies Evaluating The Reliability Of Autonomous Vehicles Based On Disengagements And Reaction Times. International Journal of Automotive Science and Technology. 2024; 8 (3): 279-287. https://doi.org/10.30939/ijastech..1480056

1. Introduction

One of the most important criteria for ensuring the good functioning of any system is its reliability. Failures in systems used for commercial purposes will lead to financial losses, and failures in systems that require security will cause security concerns. Artificial intelligence (AI) systems are gaining importance in many fields, including healthcare, education, energy, information technology, finance, transportation, and manufacturing. Improving the reliability of AI systems, which are being used in many sectors rapidly during a short period, is highly important. While a failure related to AI systems used in the production sector may lead to financial losses, failures in AI systems used in the transportation sector may cause more severe losses. For example, a failure in an AV could cause an accident. For this reason, improving AV reliability has become an inevitable need. To enhance the reliability of AVs, the Society of Automotive Engineers (SAE) has ranked AVs by introducing different autonomy levels. For this purpose, the SAE has classified AVs into six levels of automation. Here, the levels are listed from level 0, which is no automation, to level 5, which is full automation. In automation levels 0-2, the automation systems support the vehicle's control tasks, requiring human drivers to monitor the vehicle's status. However, in levels 3-5, the automatic driving systems assume full responsibility, providing no need for human control [1]. Studies on the reliability of AVs usually specify the level of automation system for which they are being studied.

Many factors affect the reliability of AVs. For this reason, several researchers have considered AV reliability a multidisciplinary issue and made inferences [2, 3]. The most important factors affecting AV reliability are identified as software and hardware defects of the system, as well as weather conditions, traffic conditions, and road surface conditions. Studies have revealed that long distances need to be traveled to test the reliability of AVs [4]. This requires a long period and regular data recording to test the reliability of the vehicles. Due to the need to test the reliability of AVs, in 2014, the California

(CA) Department of Motor Vehicles (DMV) allowed AV manufacturers to establish an AV Testing program to test their AVs. The CA DMV requires all manufacturers to test AVs on public roads and report in detail all accidents and disengagements that occur.

Disengagement events represent instances where autonomous systems are unable to operate safely due to current conditions or environmental factors. The frequency and causes of these events directly impact the safety performance and reliability of AVs. This data provides developers with the opportunity to identify weak points in the system and areas that need improvement. Consequently, through software updates and hardware enhancements, the overall reliability of the system can be increased. In situations where the AV transfers control to the human driver, the driver's reaction time is critical; a quick and appropriate response can prevent potential accidents and enhance safety. Reaction times assess the effectiveness of the interaction between human drivers and autonomous systems. Well-designed human-machine interfaces (HMIs) enable drivers to respond more quickly and effectively. HMI refers to how drivers and passengers interact with and control vehicle systems. These designs are optimized to ensure that users interact with vehicles more safely and efficiently. Hence, reaction times are a vital metric for evaluating the performance of autonomous systems and determining how these systems can work more safely and harmoniously with drivers. In this study, evaluations were made on the maturity of AVs by examining studies conducted on disengagement and reaction times, which are crucial aspects of AV reliability. A synthesis of studies providing information on the frequencies, temporal changes, reasons, and relationships among factors causing disengagements was provided and collectively assessed. The statistical methods used in evaluations of disengagements were examined and categorized to guide researchers working in this field.

The remainder of the article is organized as following. In Section 2, a brief overview of disengagements is provided along with studies conducted in this area. Section 3 presents a brief overview of reaction times and examines the studies that were conducted. Section 4 presents the results obtained from the conducted studies and provides evaluations. Finally, in Section 5, conclusions and information regarding future research are provided.

2. Disengagements

In autonomous systems, disengagement events are defined as failures that cause vehicle control to pass from the software system to the human driver. CA DMV [5] defines disengagement as “disengagement means a deactivation of the automation mode when a failure of the autonomous technology is detected or when the safe operation of the vehicle requires that the AV test driver disengages the autonomous mode and takes immediate manual control of the vehicle". Disengagements are broadly divided into two types:

1. Automatic disengagement: Automatic drive disengagement due to a detected failure or the possibility of a potential failure. In this case, the AV notifies the driver of the fault, and the driver must immediately take over driving. Failures that cause automatic disengagement can usually be caused by software problems such as hardware inconsistencies, detection errors, communication errors, incorrect sensor readings, mapping problems, and errors in data acquisition as well as weather conditions and road surface conditions [6].
2. Manual disengagement: These are disengagements where the human driver voluntarily takes control of the vehicle. Manual disengagement generally occurs because drivers need to be careful when they feel in danger in heavy traffic, in adverse weather conditions, in situations such as construction and road works in the surrounding area, or when they want to change lanes [6].

Since disengagement significantly impacts AV reliability, it has attracted the attention of many researchers, and studies have been conducted on this subject.

Dixit et al. [7] conducted a study on disengagement, accidents and reaction times in AVs between September 2014 and November 2015. Considering the relationship between drivers' trust in automated systems and their ability to manage their perceived risks effectively, researchers evaluated the correlation between monthly automatic disengagement per mile and manual disengagement per mile, and found a high correlation. Based on this, they concluded that if drivers increase the number of automatic disengagements based on their experience, the likelihood of manual disengagements will increase. They also examined the correlation between cumulative exposure to automatic disengagement and manual disengagement. They found it to be significantly larger, given that the driver's disengagement experience is based on cumulative numbers. After reviewing companies' reported reasons for disengagement, they determined that the most common cause of disengagement was system failures, followed by driver-related disengagements. Factors such as misperception of road infrastructure, incorrect lane markings, poor road conditions and other road users were the most common causes of disengagements.

SRGMs are models used to enhance the reliability of software systems. These models develop various assumptions regarding the nature of software faults, the effectiveness of fault correction, and testing patterns. Different SRGMs offer diverse approaches to understanding the occurrence and resolution of software faults, aiming to optimize software reliability [8]. Merkel [9] used software reliability growth models (SRGMs) to model disengagement data. The study used CA DMV data from 2015 to 2017. Waymo and Cruise Automation, which performed the most road tests in the given date range, were selected for analysis. To evaluate which model best fit each AV manufacturer's disengagement event data, nonlinear least squares fits were calculated for the Musa-Okumoto and Gompertz models using these data. The results showed that the Musa-Okumoto model fit the data better than the Gompertz model. In both cases, the actual data were found to be broadly consistent with the model predictions.

Lv et al. [6] analyzed disengagement events in detail from 2014 to 2015 with data from CA DMV. In their study, they analyzed data from SAE's Level 2 and Level 3 automation technologies to provide a better understanding of situations in which the driver should take over. The authors categorized the test disengaged individuals into different groups according to their attributes and investigated and compared the reasons for the disengagement in detail. The mechanisms involved and the duration of the takeover transition that occurs during disengagement were also analyzed. As a result of the study, it was determined that the most critical factors causing disengagement were found to be software problems and limitations. They have used miles per disengagement (MPD) to evaluate the maturity of autonomous technology.

Banerjee et al. [10] analyzed accident and disengagement events from the CA DMV between 2014 and 2017 to investigate the causes of AV failures. In the study, AVs, expressed as level 3 by SAE, were discussed. The researchers conducted a study based on an end-to-end data collection, processing and analysis pipeline. Data from 12 autonomous vehicle manufacturers, obtained from the CA DMV, were collected and analyzed using natural language processing (NLP) techniques. The NLP method can quickly analyze large amounts of text data to extract important information. In this study, a "Faulire Dictionary" was created to assign fault labels to the reasons for disengagement from raw disengagement reports. However, the lack of a specific data format in manufacturers' disengagement and accident reports can lead to systematic errors in the NLP framework. To minimize this, researchers have labeled uncertain cases as "unknown" in the manually validated dictionaries. Based on available data, the number of disengagements per autonomous mile (DPM) and the cumulative number of disengagements were used to evaluate AVs. They found that despite covering millions of miles, all manufacturers involved in the study, including Waymo, were still in the early stages of development. It was determined that the primary reasons for disengagements in AVs were inconsistencies related to machine learning in perception, control, and decision-making. Additionally, it was noted that AV drivers need to be as observant as drivers operating non-autonomous vehicles.

Favarò et al. [12] analyzed triggers and contributing factors, such as the average distance traveled before failure and disengagement reporting trends, using data from the CA DMV covering September 2014 to January 2017. Their analysis identified deficiencies in current regulations and provided suggestions for improvement in the existing drafts. The results revealed that system failures played a dominant role in AV disengagements, with hardware-related failures being more prevalent than software-related ones. The researchers also plotted the frequency of disengagements (per total miles driven) as a function of cumulative miles or time, yielding significant insights. A linear regression analysis showed the relationship between cumulative reported autonomous miles and cumulative disengagements, emphasizing the importance of experiences gained during the development of AV technology. A critical step in designing safe and effective AVs involves training the AI and machine learning-based algorithms that act as the vehicle's "brain.". As AVs encounter various training scenarios, they improve their ability to handle new situations by learning from mistakes. Furthermore, the study noted that disengagement of AVs does not always result in an accident.

Studies conducted before Zhao et al. [13] did not take into account any information about the reliability of the AV before the road test. Therefore, the researchers proposed a new Conservative Bayesian Inference (CBI) method that will allow for the prediction of future AV disengagements from past AV disengagement data. For this purpose, they used SRGMs. Researchers have shown that SRGM can be an effective test planning tool when combined with accuracy assessment and recalibration techniques. SRGMs were applied to 51 months of test data obtained from Waymo. In the study, recalibration, comparison and visualization techniques were applied with 8 SRGMs using a state-of-the-art toolkit called PETERS. Using previous interfailure mileage data, and the most reliable SRGM was selected to estimate the median miles to the next disengagement (MMTD). Methods introduced by Brocklehurst et al. [14] and Brocklehurst and Littlewood [15] have been used to estimate historical disengagement trends with SRGMs. Like previous studies, this study notes that SRGMs may not always provide the "best" estimates. Therefore, it is important to evaluate forecast accuracy continually. It is known that even if there is a finding that one SRGM performs better than others on one set of predictions, this may change in subsequent observations. The study also shows that systemic shortcomings in past forecasts can be addressed by using recalibration techniques to improve the performance of these models.

Wang and Li [16] examined disengagement data obtained from the CA DMV for the years 2016-2017 using multiple statistical modeling approaches that involve statistical modeling and classification tree. In the study, AVs classified as Level 3 and Level 4 by the SAE were examined. The research aims to identify factors influencing human drivers' quick response to AV disengagement and quantitatively investigate the underlying reasons for AV disengagement. In the study, ordinal logistic regression was applied to determine the reason for disengagement using various factors, such as the number of camera, LIDAR, and radar sensors, as categorical independent variables. The time required to take over driving was divided into two categories: "within 0.5 s" and "over 0.5 s" and modeled using binary logistic regression. Additionally, using the Classification and Regression Tree (CART) model, a classification tree was created if the dependent variable was categorical, and a regression tree was created if the dependent variable was numerical. This model identified the relationship between factors related to disengagements and the time required to take over driving. Classification trees have two important components: the "root node" and the "leaf nodes" [17]. Compared to other machine learning techniques like Random Forest, the CART model has the advantage of quantitatively analyzing the leaf nodes affecting nominal dependent variables. According to the findings of the research, it was concluded that to reduce interference in autonomous driving systems, 5 or more radar sensors should be installed in AVs, the number of LiDAR sensors should be 3 or 4, and the number of cameras can be customized according to the preference of each AV manufacturer. To address the issue of extended takeover time, it was found that drivers can usually take over within 0.5 seconds on local roads, but the disengagement of AVs on highways due to detection or control issues extends the takeover time.

Boggs et al. [18] examined in detail CA DMV disengagement data from September 2014 through November 2018 with five W questions (i.e., who, what, when, where, and why). The researchers stated that the disengagement initiator is associated with factors such as whether it is an automated driving system

(ADS) or a human operator, its location, the reason for disengagement, and the maturity of the ADS test. They determined this relationship using a random parameter binary logit model. Disengagements in AVs can be caused by various factors, and not all of these potential factors can be captured from the available data. This situation is referred to as "unobserved heterogeneity" in transportation studies [19, 20]. Considering the potential for inconsistent and biased parameter estimates, the study employs the random parameter logit model, which is frequently used in the literature, to account for unobserved heterogeneity. According to the study's findings, streets and roads were identified as locations where ADS was less likely to disengage compared to highways and freeways. This is attributed to greater interactions with other vehicles on streets and roads, and the complexity of factors such as intersections and vehicle entries compared to highways and freeways. Consequently, it was found that the probability of detecting unexpected events on streets and roads was lower compared to situations where ADS actively monitored individuals. Additionally, the results showed that due to hardware and software discrepancies, planning diversity, environmental factors, and the influence of other road users, the probability of AVs becoming disengaged is higher than control non-compliance.

Table 1. Summary of Studies on Disengagement Events for AV Reliability: References, Data Types, Companies, Periods, and Methods

References	Data Type	Company	Period	Method
(Dixit et al. 2016)	CA DMV	Bosch, Delphi Automotive, Google (now Waymo), Mercedes-Benz, Nissan, Volkswagen and Tesla	September 2014-November 2015	Disengagement exposure per autonomous miles, descriptive statistics (correlation between the monthly automatic disengagements/mile and manual disengagements/mile, correlation between the cumulative exposure of automated disengagements and manual disengagements).
(Merkel 2018)	CA DMV	Waymo (former Google), Cruise	2015-2017	Software Reliability Growth Models, nonlinear least squares fit.
(Lv et al. 2018)	CA DMV	Bosch, Delphi Automotive, Google (now Waymo), Mercedes-Benz, Nissan, Volkswagen and Tesla	2014-2015	Descriptive statistics (autonomous miles per disengagement).
(Banerjee et al. 2018)	CA DMV	Bosch, Delphi Automotive, Google (now Waymo), Nissan, Mercedes-Benz, Tesla, BMW, GM, Ford, Honda, Uber and Volkswagen	September 2014-November 2016 (included in the DMV's data releases for 2016 and 2017)	Natural language processing, the end-to-end data collection, processing, and analysis pipeline, Systems-Theoretic Process Analysis, descriptive statistics (autonomous miles per disengagement).
(Favarò et al. 2018)	CA DMV	Bosch, Delphi Automotive, Google (now Waymo), Nissan, Mercedes-Benz, Tesla Motors, BMW, GM, Ford, Honda and Volkswagen group of America	September 2014-January 2017	Regression analysis.
(Zhao et al. 2019)	CA DMV	Waymo (former Google)	2014 - 2019 (over 51 months)	Software Reliability Growth Models, a new variant of Conservative Bayesian Inference.
(Wang and Li, 2019)	CA DMV	The companies operating within the given date range	August 2016-November 2017	Ordinal and binary logistic regression, classification and regression tree model.
(Boggs et al. 2020)	CA DMV	The companies operating within the given date range	September 2014-November 2018	Random parameter binary logit model.
(Khattak et al. 2021)	CA DMV	The companies operating within the given date range	2014-2018	Text mining, nested logit model and endogenous switching model.
(Zhang et al. 2022)	CA DMV	The companies operating within the given date range	2014-2020	Natural language processing, deep transfer learning.
(Min et al. 2022)	CA DMV	Waymo (former Google), Cruise, Pony AI, Zoox	December 2017-November2019	Software Reliability Growth Models, Spline models.

These findings are an intuitive assessment of individuals' inability to monitor the data processing of Autonomous Driving Systems (ADS) and detect any inconsistencies in the system. Finally, it was concluded that with the maturity of the ADS system, there would be a marginal increase in disengagements initiated by ADS as operators begin to trust and rely on the AV system when facing risky situations.

Using CA DMV disengagement data published between 2014 and 2020, Zhang et al. [21] developed a scalable pipeline incorporating NLP and deep transfer learning methods. In the study, the NLP pipeline utilized deep transfer learning to improve the extraction of cause-and-effect relationships in AV disengagements (AVD). This method enhances the learning process by transferring knowledge from previously learned natural language understanding tasks to new tasks. Within the scope of this study, the classification, visualization, and analysis of disengagement data were conducted using statistical tests, revealing significant relationships between trends in AV testing, the frequency and origins of disengagements, and their impacts. From this study, it was found that manufacturers tend to intensively test AVs in Spring and/or Winter months, with test drivers initiating over 80% of disengagements. More than 75% of disablings are caused by errors in perception, localization, mapping, planning and control of the AV system. This highlights that there is a significant relationship between the person initiating the AVD and the error category.

Khattak et al. [22] aimed to explain the relationship between disengagements and accidents in complex traffic environments, using CA DMV data between 2014 and 2018. To this end, they analyzed accidents and disengagements, investigating three different situations: (1) disengagement with an accident, (2) disengagement without an accident, and (3) no disengagement when there was an accident, using a nested logit model. Additionally, endogenous regime-switching models were employed to account for endogeneity effects in determining the differences between disengagements and accidents. Nested Logit and Endogenous Switching Regime models were used to obtain more accurate predictions by accounting for the correlation between choices and the effects of endogenous selection. The Nested Logit model considers correlations between alternatives through a nested structure, while the Endogenous Switching Regime model accounts for differences between collisions and non-collision disengagements, as well as endogenous effects. They conducted this process using text mining on the data. The results showed that disengagements in AVs do not always cause accidents, that factors related to AV systems (such as software errors) increase the probability of disengagement, and that factors related to the driver's decision-making process increase the likelihood of disengaged vehicles resulting in an accident. The findings also show that disconnection does not always have negative consequences, but it is critical for safe functioning. Moreover, it has been observed that as technology matures, AVs disengage less frequently.

Min et al. [23] evaluated the reliability of AI systems using data on recurrent disengagement events reported by Waymo, Cruise, Pony AI, and Zoox, which conducted peak road tests between December 1, 2017 and November 30, 2019. For this purpose, they used SRGMs, which are parametric models, and the I-spline model, which is a nonparametric model. In addition to traditional parametric models, they proposed a new nonparametric model based on monotonic splines for software reliability.

By selecting the best models, they made inferences to quantify uncertainty and test heterogeneity in the event process. The results show that the proposed spline model is flexible in explaining recurring event data from four AV manufacturers and that parametric models are adequate for most manufacturers' data. Additionally, from data analysis, it was determined that overall AV reliability increased over the 2-year study period. A summary of studies on disengagement events for the reliability of AVs is given in Table 1.

3. Reaction times

As a result of a failure of an AV, the driver must take control of the vehicle as soon as the autonomous system is disabled. The automation system can request a takeover action or initiate it manually by the driver/operator. The time it takes when the driver is alerted to a technology failure and assumes manual control of the vehicle is called reaction time or response time. Reaction time determines how quickly a person reacts when faced with risk and is a critical factor in avoiding accidents. The disengagement process, together with the takeover process, is key and greatly affects the safety and comfort of automated vehicles. For this reason, reaction times have been studied in the literature. Research on reaction times during the handover process in AVs is categorized into two main areas. The first category encompasses studies conducted using simulators. In these studies, drivers' reaction times to various scenarios (e.g., traffic density, emotional state, non-driving related tasks) are evaluated. Simulation studies provide the opportunity to collect detailed data under controlled and repeatable conditions, allowing for the analysis of different scenarios' effects. Several studies in the literature conducted using simulators are highlighted here. Du et al. [24] examined the impact of traffic density on reaction time under different cognitive load conditions. Their study found that reaction time performance decreased under high cognitive load and high traffic density, whereas under low cognitive load, high traffic density improved reaction time performance. Gold et al. [25] investigated reaction times at three different traffic density levels (zero, 10, and 20 vehicles per kilometer). They found that traffic density had a significantly negative impact on reaction time performance. Du et al. [26] also studied the effects of emotional state and arousal level on reaction time, discovering that positive emotions enhanced takeover performance, while loud alerts provided no advantage. One of the most anticipated features of AVs is the ability for drivers to engage in non-driving related activities (e.g., messaging, emailing, watching videos) while driving. Consequently, takeover times during non-driving related tasks have become a subject of interest for researchers, leading to numerous studies in this area [27, 28, 29, 30]. Studies in the first category utilize statistical methods such as Analysis of Variance (ANOVA) to compare means between different groups and determine whether there are significant differences. MANOVA (Multivariate Analysis of Variance) allows for the simultaneous analysis of multiple dependent variables, providing a more comprehensive assessment of the effects of age and activity level on the frequency and duration of non-driving activities. These approaches are crucial for understanding the potential impact of non-driving activities on driving safety. The second category focuses on reaction times obtained from real driving conditions. These studies examine drivers' reactions in real-world driving scenarios and in real-time. These studies from the literature are presented below.

Dixit et al. [7] found that when the vehicle is disengaged, reaction times across different companies averaged 0.83 seconds and had a stable distribution. However, differences in reaction times may be observed depending on factors such as disengagement type, route type, and trip length. It has been found that the lack of confidence caused by automatic disengagement increases the likelihood that the driver will take manual control of the vehicle. When the relationship between reaction times and total monthly AV kilometers was examined, it was found that as vehicle kilometers increased, reaction times also increased. This shows that as vehicle mileage increases, the confidence level also increases.

Lv et al. [6] examined manufacturers' takeover transition mechanisms and the duration of the takeover transition in a takeover process, with data obtained from CA DMV between 2014 and 2015. According to their findings, this usually occurs within 1 second. However, researchers have emphasized that the takeover transition is not always a simple task that can be completed quickly and smoothly. During testing of automated vehicles, it is assumed that test drivers are optimized for all scenarios and ready to take over control of the vehicle. However, in real life, all drivers have different training levels and concentration levels. Therefore, researchers emphasize that accurately detecting driver behavior and attention levels is a significant challenge for effectively designing human-machine interfaces.

Banerjee et al. [9] emphasized that how quickly drivers react in case of failure is important to reduce the risk of accidents. In their study, they examined the distribution of reaction times of test drivers from different manufacturers and determined that this distribution was long-tailed and compatible with the exponential Weibull distribution. They also found the average reaction time for all manufacturers to be 0.85 seconds. This result is consistent with previous findings by Dixit et al. [7].

Hecker et al. [31] developed the concept of “scene driveability”, a camera-based driving model, and trained it with real driving datasets. In this study, a new learning method based on recurrent neural networks was used to determine the suitability of a particular driving scene for a particular autonomous driving method. The study complements existing ADAS and driver monitoring techniques by adding to fully automated cars the ability to predict automation failures and provide timely alerts to the human driver. In this way, human drivers can be warned in real time, increasing the overall safety of the autonomous driving model and allowing better human-vehicle cooperation.

4. Discussion and concluding remarks

Important results and findings from studies on the reliability of AVs are discussed in this section. Researchers have used real data from the CA DMV in their studies to explain the occurrence of disengagement, as values from real datasets provide us with more realistic evidence for the possible causes of disengagement. Although the studies include Level 4 and below among the levels determined by SAE, they are mainly focused on Level 3. To examine possible causes of disengagements, researchers examined the frequent disengagemets per autonomous mile driven. This approach helps estimate the frequency of disengagemet and the average distance traveled before disengagemet. It is also important to better understand the relationship between distance traveled and disengagemets [6, 7, 10, 12]. According to these studies, there is a strong correlation between autonomous miles traveled and disengagemets, and with the development of AV technology, AVs learn from their mistakes with trained data and become more successful at dealing with new situations.

Established in 2014 within the CA DMV, initially disengagement reports were provided in bulk on a monthly basis until November 1, 2017. However, later on, the content of the reports was expanded, enriching them with detailed data such as the date of disengagement, Vehicle Identification Number (VIN), the vehicle's autonomous driving capability, the location where the disengagement occurred, road conditions, and a detailed description. Additionally, a significant increase in test drives has been observed since 2019. The increase in data volume and information has led researchers to resort to methods such as NLP, data mining, and text mining to derive deeper insights from this information [10, 21, 22]. Furthermore, researchers have frequently used regression and classification methods to classify disengagement data and identify factors influencing and triggering disengagements and their relationships [12, 16, 18, 22].

Considering that categorizing disengagemets is important for determining the causes of disengagement, researchers have also examined the relationships between types of disengagement. Dixit et al. [7] examined both the relationship between automatic disengagement and manual disengagement per mile per month and the cumulative exposure to automatic disengagement and manual disengagement, as a driver's experience with disengagement is based on cumulative numbers, and found high associations in both cases. They also found that system failure was the most common type of disengagement, involving hardware and software problems, followed by driver-initiated disengagement. Lv et al. [6] found that software problems and limitations are the most important factors that lead to disengagement. Banerjee et al. [10] reported that, in terms of the reasons for the occurrence of disengagemets, machine learning-related errors, especially those related to the sensing system, were the dominant cause of disengagemets for most manufacturers, while the second largest contributor to disengagemets was machine learning related to the control and decision framework. The computing system, i.e., hardware issues and software issues, accounted for approximately 33.6% of the total reported disconnections. According to the results obtained, Favarò et al. [12] found that system failures have a dominant role in the failure of AVs. According to their findings, they also stated that software-related failures play a greater role in disengagemets than hardware-related failures. Boggs et al. [18] found that ADS less frequently disengages streets and roads compared to highways and freeways. Additionally, they identified that the probability of disengagement initiation is higher due to hardware and software discrepancies, planning discrepancies, environmental factors, and interactions with other road users compared to control discrepancies. In a study conducted by Zhang et al. [21], it was concluded that manufacturers tend to extensively test AVs during the spring and/or winter months, with over 80% of disengagements initiated by test drivers. However, according to the findings, it was revealed that more than 75% of the disengagements were caused by errors in perception, mapping, localization, planning, and control of the AV system. Additionally, a significant relationship was found between the person who initiated the disengagement of AVs and the category of the cause. Khattak et al. [22] found that factors related to AV systems increase the probability of non-collision failure in cases such as software errors; observed that factors related to the driver's decision-making process increase the likelihood of a disengagement resulting in a crash.

Several researchers have used SRGMs to determine the most appropriate models for disengagement data [9, 13, 23]. Merkel [9] found that the Musa-Okumoto model, an SRGM model, is more suitable for the data than the Gompertz model is. Again, according to the results obtained, the disengagement events reported in two large AV test programs can be accurately adapted to standard SRGMs. The actual data were broadly consistent with the model predictions in both cases. Zhao et al. [13] used SRGMs to demonstrate how past AV disengagement data can be used to predict future disengagement with a novel conservative Bayesian inference (CBI) method. As with previous work on SRGMs, this study emphasized the importance of continually assessing prediction accuracy, as various applications have shown that a given SRGM should not always be expected to yield the "best" predictions. Min et al. [23] used SRGMs, which are parametric models, and the I-spline model, which is a nonparametric model, to explain the disengagement event processes. Based on the I-spline model, they found that the most appropriate model can be selected by measuring its uncertainty and testing the heterogeneity in the event process. Parametric models and spline models are recommended as complementary tools in modeling and inference processes.

Finally, reaction times were discussed based on both real data and results obtained through simulators. The findings from real data are summarized as follows. Dixit et al. [7] found that in the event of vehicle disengagement, the reaction times to take control of the vehicle have a stable distribution with an average of 0.83 s across different companies. However, it was observed that there might be differences in reaction times depending on the type of disengagement, type of roadway, and km traveled. According to the study by Lv et al. [6], most of the average values of the reported takeover times were within 1 second. In their study, Banerjee et al. [10] showed the distribution of test drivers' reaction times among all manufacturers and determined an average reaction time of 0.85 seconds for all test vehicle drivers. In addition, researchers emphasized that human drivers in AVs should be alert and sensitive to the environment.

After disengagements in AVs, the performance of human drivers in taking over control depends on factors such as traffic density, road conditions, and weather conditions, as well as personal characteristics and psychological factors like cognitive load and emotional state. Additionally, one of the primary reasons for using an AV in autonomous mode is to perform non-driving related tasks, such as messaging, emailing, or watching videos. Consequently, researchers have conducted numerous studies evaluating all these situations, with examples provided in the reaction times section. Real driving data has higher ecological validity because it reflects drivers' natural behaviors and can reveal the effects of factors that cannot be observed in simulator studies. These two approaches are complementary, helping to develop a comprehensive understanding of the safety and effectiveness of transitioning from autonomous to manual driving.

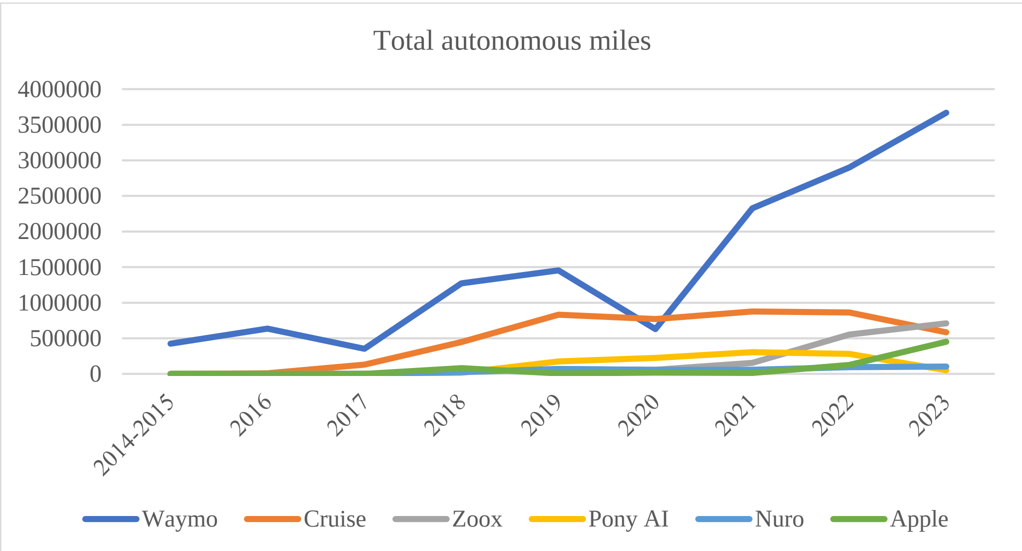

Fig. 1. Total number of autonomous miles driven by manufacturers in California between 2014 and November 2023.

A review of recent studies shows that since 2014, when the CA DMV allowed AV manufacturers to test automated systems, studies in this field have increased rapidly. While the reliability of autonomous vehicles was generally investigated using descriptive statistics until 2018, since then, there has been an increasing emphasis on machine learning-based methods. In their studies, researchers often evaluated the development, maturity, and stability of autonomous vehicle technology, particularly about its reliability. It has been stated that comprehensive road tests are required to detect these [4]. Additionally, miles per disengagement is frequently calculated to determine the stability and maturity of AV technology [7, 6, 10, 12]. To address these questions, data on the total autonomous miles driven since the first registrations of manufacturers with the CA DMV have been provided. Figure 1 presents the total autonomous miles driven by the six manufacturers conducting the most road tests from 2014 to the end of 2023 [32]. These manufacturers are Waymo, Cruise, Zoox, Pony AI, Nuro, and Apple. Among these manufacturers, only Waymo has reported road reports to the CA DMV since 2014. Cruise began reporting in 2016, Zoox in 2017, and Nuro and Apple from 2018 onwards. Waymo has comprehensive data since the inception of registration with the CA DMV and has the most extensive road tests.

5. Conclusion

This review study provides an evaluation of disengagements and reaction times, which are crucial topics in AV reliability research, through the analysis of statistical methods employed. Studies on the reliability of AV technology were predominantly conducted before 2019. However, as illustrated in Figure 1, there has been a significant increase in road tests, especially since 2020. The substantial rise in road tests and the increased detail in reports provided by the CA DMV offer much more comprehensive information regarding disengagements, accidents, and reaction times. Utilizing this data with appropriate statistical methods can yield numerous valuable insights. Classical statistical tests, such as regression and variance analysis, can be employed to determine the impact of factors like road conditions, driver behaviors, and vehicle characteristics on the risk of disengagement or accidents. These analyses allow for a better understanding of the causes and risk factors of accidents and disengagements. Machine learning techniques, particularly decision trees, support vector machines, and deep learning algorithms, can be used to develop disengagement and accident prediction models and to foresee high-risk situations. While large datasets can be used with machine learning algorithms to create disengagement and accident prediction models, data mining techniques can uncover underlying patterns and hidden relationships in the data. This information can be used to develop new traffic regulations, evaluate the effectiveness of current policies, and identify high-risk areas. Collecting detailed data and conducting appropriate statistical analyses will facilitate the identification and improvement of the shortcomings of AVs, leading to safer driving experiences.

Conflict of Interest Statement

The authors declare that they had no agreement or financial involvement with anyone, organization or institution in the course of this research work. Therefore, there is no conflict of interest associated with any part of this paper or the publication of this manuscript.

CRediT Author Statement

Ferhan Baş Kaman: Data curation, Formal analysis, Investigation, Conceptualization, Writing, Review and Editing.
Hülya Olmuş: Supervision, Investigation, Writing, Review and Editing.

References

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

[2] Koopman, P., Wagner, M. Autonomous vehicle safety: An interdisciplinary challenge. IEEE Intell. Transp. Syst. Mag.. 2017, 9(1), 90–96. https://doi.org/10.1109/MITS.2016.2583491

[3] Burton, S., Habli, I., Lawton, T., McDermid, J.,Morgan, P., Porter, Z. Mind the gaps: assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. Artificial Intelligence. 2020, 279, 103201. https://doi.org/10.1016/j.artint.2019.103201

[4] Kalra, N., Paddock, S.M. Driving to safety: how many miles of driving would it take to demonstrate autonomous vehicle reliability? Transportation Research Part A: Policy and Practice. 2016, 94, 182–193. https://doi.org/10.1016/j.tra.2016.09.010

[5] California Department of Motor Vehicles (CA DMV). Article 3.7– Autonomous Vehicles. Title 13, Division 1, Par. 227. 2016. https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/testing

[6] Lv, C., Cao, D., Zhao, Y., Auger, D.J., Sullman, M., Wang, Dutka, H. L. M., Skrypchuk, L., Mouzakitis, A. Analysis of autopilot disengagements occurring during autonomous vehicle testing. IEEE/CAA Journal of Automatica Sinica. 2018, 5(1), 58–68. http://dx.doi.org/10.1109/JAS.2017.7510745

[7] Dixit, V.V., Chand, S., Nair, D.J. Autonomous vehicles: disengagements, accidents and reaction times. PLoS ONE. 2016, 11(12): e0168054. https://doi.org/10.1371/journal.pone.0168054

[8] Wood, A. Software Reliability Growth Models. Tandem, Technical Report 96.1, Tandem Computers, 1996; Cupartino, CA.

[9] Merkel, R. Software reliability growth models predict autonomous vehicle disengagement events. arXiv: 1812.08901.2018. https://doi.org/10.48550/arXiv.1812.08901

[10] Banerjee, S.S., Jha, S., Cyriac, J., Kalbarczyk, Z.T., Iyer, R.K. Hands off the wheel in autonomous vehicles? A systems perspective on over a million miles of field data. In 2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). 2018, 586–597. https://doi.org/10.1109/DSN.2018.00066

[11] Leveson, N. Engineering a safer world: Systems thinking applied to safety. MIT press; 2011.

[12] Favarò, F., Eurich, S., Nader, N. Autonomous vehicles disengagements: trends, triggers, and regulatory limitations. Accident Analysis & Prevention. 2018, 110, 136–148. https://doi.org/10.1016/j.aap.2017.11.001

[13] Zhao, X., Robu,V., Flynn,D., Salako, K., Strigini, L. Assessing the safety and reliability of autonomous vehicles from road testing. In 2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE). 2019, 13–23. https://doi.org/10.1109/ISSRE.2019.00012

[14] Brocklehurst, S., Chan, P. Y., Littlewood, B., Snell, J. (1990). Recalibrating software reliability models. IEEE Transactions on

Software Engineering. 1990, 16(4), 458–470. http://dx.doi.org/10.1109/32.54297

[15] Brocklehurst S., Littlewood, B. Techniques for prediction analysis and recalibration. in Handbook of Software Reliability Eng., M. Lyu, Ed. McGraw-Hill & IEEE Computer Society Press; 1996.

[16] Wang, S. and Li, Z. Exploring causes and effects of automated vehicle disengagement using statistical modeling and classification tree based on field test data. Accident Anal. Prevention. 2019, 129, 44-54. https://doi.org/10.1016/j.aap.2019.04.015

[17] Weng, J., Meng, Q. Decision tree-based model for estimation of work zone capacity. Transportation Research Record: Journal of the Transportation Research Board. 2011, 2257, 40–50. http://dx.doi.org/10.3141/2257-05

[18] Boggs, A. M., Arvin, R., Khattak, A.J. Exploring the who, what, when, where, and why of automated vehicle disengagements. Accident Anal. Prevention. 2020, 136, 105406. https://doi.org/10.1016/j.aap.2019.105406

[19] Wali, B., Khattak, A.J., Khattak, A.J. A heterogeneity based case-control analysis of motorcyclist's injury crashes: evidence from motorcycle crash causation study. Accident Analysis & Prevention. 2018, 119, 202–214. http://dx.doi.org/10.1016/j.aap.2018.07.024

[20] Azimi, G., Asgari, H., Rahimi, A., Jin, X.. Investigation of heterogeneity in severity analysis for large truck crashes. 98th Annual Meeting of the Transportation Research Board. 2019, Washington, D.C. United States.

[21] Zhang, Y., X. J. Yang, X. J., Zhou, F. Disengagement Cause-and-Effect Relationships Extraction Using an NLP Pipeline. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11), 21430-21439, https://doi.org/10.1109/TITS.2022.3186248

[22] Khattak, Z. H., Fontaine, M.D., Smith, B. L. Exploratory Investigation of Disengagements and Crashes in Autonomous Vehicles Under Mixed Traffic: An Endogenous Switching Regime Framework. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(12), 7485-7495. https://doi.org/10.1109/TITS.2020.3003527

[23] Min, J., Hong, Y., King, C. B., Meeker W. Q. Reliability analysis of artificial intelligence systems using recurrent events data from autonomous vehicles. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2022, 1–30. http://dx.doi.org/10.1111/rssc.12564

[24] Du, N., Kim, J., Zhou, F., Pulver, E., Tilbury, D.M., Robert, L.P., Pradhan, A.K., Yang, X.J. Evaluating effects of cognitive load, takeover request lead time, and traffic density on drivers' takeover performance in conditionally automated driving. In Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications; 2020; Washington DC, USA. http://dx.doi.org/10.1145/3409120

[25] Gold, C., Körber, M., Lechner, D., Bengler, K. Taking over control from highly automated vehicles in complex traffic situations: The role of traffic density. Human Factors: The-Journal of the Human Factors and Ergonomics Society. 2016, 58(4), 642–652. http://dx.doi.org/10.1177/0018720816634226

[26] Du, N., Zhou, F., Pulver, E. M., Tilbury, D.M., Robert, L.P., Pradhan, A.K., Yang, X.J. Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving. Transportation Research Part C: Emerging Technologies. 2020, 112, 78-87. https://doi.org/10.1016/j.trc.2020.01.006

[27] Clark, H., Feng, J. Age differences in the takeover of vehicle control and engagement in non-driving-related activities in simulated driving with conditional automation. Accident Analysis & Prevention. 2017, 106, 468–479. http://dx.doi.org/10.1016/j.aap.2016.08.027

[28] Wandtner, B., Schömig, N., Schmidt, G. Effects of non-driving related task modalities on takeover performance in highly automated driving. Human Factors: The-Journal of the Human Factors and Ergonomics Society. 2018, 60(6), 870–881. https://doi.org/10.1177/0018720818768199

[29] Dogan, E., Honnêt, V., Masfrand, S., Guillaume, A. Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving. Transportation Research Part F: Traffic Psychology and Behavior. 2019, 62, 494-504. https://doi.org/10.1016/j.trf.2019.02.010

[30] Hu, W., Zhang, T., Zhang, Y., Chan, A.H.S. Non-driving-related tasks and drivers' takeover time: A meta-analysis. Transportation Research Part F: Traffic Psychology and Behavior. 2024, 103, 623-637. https://doi.org/10.1016/j.trf.2024.05.012

[31] Hecker, S., Dai, D., Van Gool, L. Failure prediction for autonomous driving. In 2018 IEEE Intelligent Vehicles Symposium (IV). 2018. https://doi.org/10.48550/arXiv.1805.01811

[32] California Department of Motor Vehicles (CA DMV). Disengagement reports. 2024.