Data Authorisation and Validation in Autonomous Vehicles: A Critical Review *

Reem Alhabib and Poonam Yadav

Computer Science, University of York.

*Corresponding author(s). E-mail(s): reem.alhabib@york.ac.uk; poonam.yadav@york.ac.uk;

Abstract

Autonomous Vehicles (AVs) are becoming increasingly prevalent due to their potential to improve road safety and reduce environmental impact. These vehicles rely on Automated Driving Systems (ADS), which integrate multiple sensors and actuators. While some AVs operate with minimal human intervention, fully autonomous systems eliminate the need for human control entirely. Despite advances in AV technologies, secure and trustworthy data management remains a significant challenge.

This survey focuses on two relatively underexplored aspects in AV environments: data authorisation and validation. It examines the key related challenges and reviews existing solutions. The findings highlight critical gaps in current approaches and suggest future research directions to enhance AV data authorisation and validation.

Keywords: Autonomous Vehicle, Automated Driving System, Data Authorisation, Data Validation

This is the author version

1 Introduction

AVs are an innovation in the automotive sector as they provide safer, more effective, and ecologically friendly transportation options [1, 2]. In addition, they have the potential to significantly contribute to economic growth through various aspects. For example, they can substantially decrease the number of traffic accidents, leading to considerable economic savings [3]. Moreover, their impact on land use is notable, as

repurposing parking areas for the real estate industry can increase land value by 5% [4, 5]. Accordingly, in recent years, prominent automotive manufacturers have invested substantially in a diverse range of AV technologies, amounting to billions of dollars. Several sources estimate that the AV market share will take between 15 and 20 years to reach 25 percent globally [6]. In line with this trend and industrial development, governments worldwide provide guidance that allows and encourages on-road AV trials. For example, the UK published a code of practice (2019) that specifies certain road trials that include various levels of automation [7]. Subsequently, in early 2022, the government changed the Highway Code to ensure the first self-driving vehicles are introduced safely on the roads [8].

The Automated Driving System (ADS) is an integrated vehicle system that utilises various in-vehicle technologies and sensors to navigate autonomously from a starting point to a predefined destination. It comprises multiple control units that design a system to complete all driving tasks without human intervention. Cameras, GPS, and other sensors are connected to exchange data to facilitate independent driving decisions. These systems operate in a dynamic environment that demands real-time, rapid data feeding from various sources, including external roadsides and other vehicles, to the onboard sensors, which need to make continuous control decisions. Ensuring trust in automated driving systems relies heavily on the integrity and reliability of the surrounding data ecosystem. Thus, optimising data use is essential to improve the functionality of autonomous car systems. However, data collection, generation, processing, and storage challenges present critical research areas. Issues such as data privacy, integrity, and accessibility must be addressed to ensure reliable decision-making in real-time. As the complexity of these data interactions increases, innovative solutions are required to manage and safeguard the vast amounts of information generated by ADS.

While previous surveys have explored data security and privacy in AV systems, they often lack a dedicated focus on authorisation and validation mechanisms or discuss these aspects only at specific stages rather than across the entire data lifecycle. This paper critically examines the data aspects of AVs and provides an overview of ADS technology. In particular, it offers important insights into the definition, structure, data flow, ownership dynamics, and difficulties associated with AVs. Its systematic method strengthens credibility and advances knowledge of the consequences of AV data management.

1.1 Contribution

This paper aims to provide a comprehensive understanding of data and information flow in AV with a particular focus on authorisation and validation challenges across different data processing stages. Unlike previous surveys that primarily examine security, privacy, or general validation frameworks, our study provides a stage-specific analysis of data authorisation and validation challenges. As shown in Table 1, existing surveys often focus on specific aspects, such as security risks, blockchain-based access control, or legal considerations, while our work systematically categorises these challenges based on the data flow stages: collection, transmission, processing, actuation and storage.

1.2 Paper Structure

In this paper, section 2 provides a brief overview of ADS, describing the information required to understand the structure and installation of these vehicles and all related technologies, including an overview of ADS's potential benefits and costs. It also introduces key data authorisation and validation concepts, explaining their roles in ensuring secure and accurate data handling. Section 3 discusses various data management issues, with each subsection addressing the requirements, current problems, and gaps and presenting solutions for each data lifecycle stage. Section 4 reviews other related concerns, such as security, privacy, and ethics. Section 5 considers open questions and future work.

1.3 Methodology

This section describes, as in Figure 1, the methodical review strategy used to make the search for and choosing a review strategy transparent and clear. The methodology of this survey follows a systematic approach aimed at understanding authorisation and validation within the data lifecycle of autonomous vehicles. It begins with a focus on identifying relevant studies through specific inclusion and exclusion criteria, ensuring that only those addressing data-specific challenges are considered. In addition, selected studies were mapped to various lifecycle stages (such as data sources, edge computing, and cloud storage) to evaluate their impact on authorisation and validation solutions. A systematic search was conducted in major academic databases using targeted keywords to gather relevant literature.

Table 1: Comparison of Existing Surveys. The table presents a comparison of existing surveys with respect to their coverage of data validation and authorisation stages.

Survey	**Focus**	**Validation**	**Validation Stage**	**Authorisation**	**Authorisation Stage**
[9]	Security and privacy in AVs	✗	Not discussed	Partly	Data transmission, data storage
[10]	Privacy and security of Autonomous Connected Vehicles	Partly	Data exchange and processing	Partly	Data exchange and communication
[11]	Data security in autonomous driving	✗	Not covered	Partly	Communication

(Continued)

Survey	Focus	Validation	Validation Stage	Authorisation	Authorisation Stage
[12]	Data security in autonomous driving	✗	Not covered	Partly	Data collection and exchanging
[13]	Validation frameworks for AVs	Partly	System-level validation, testing	✗	Not covered
[14]	Explanations in automated driving systems	Partly	Model validation, decision assessment	✗	Not covered
[15]	Software V&V in AVs	Partly	Training and testing	✗	Not covered
This Work	Data validation and authorisation in AVs	✓	Covers all stages	✓	Covers all stages

2 Background

This section provides a comprehensive overview of the technological foundation of ADS in AVs, including the core components, communication frameworks, and main definitions. It provides foundational definitions to frame the current landscape of autonomous vehicle (AV) technologies.

2.1 Automated Driving System (ADS)

Continuous innovations have shaped the development of autonomous vehicles (AVs). Figure 2 illustrates the evolution of AVs, highlighting key technological and regulatory milestones from the 1950s to 2025. It begins with the introduction of cruise control in 1958 and includes key milestones such as the integration of road-recognition cameras in 1977.

Driving Assistance Systems (DAS) have played a foundational role in the evolution of AVs, providing critical support for vehicle control and safety. The first generation of DAS utilised sensors that assessed a vehicle's internal condition, primarily focused on safety and stability. In the 1980s, DAS included systems like Traction Control Systems (TCS) and Anti-lock Braking Systems (ABS), which aimed to improve dynamic vehicle stability [16]. In 1995, Electronic Stability Control (ESC) was introduced to

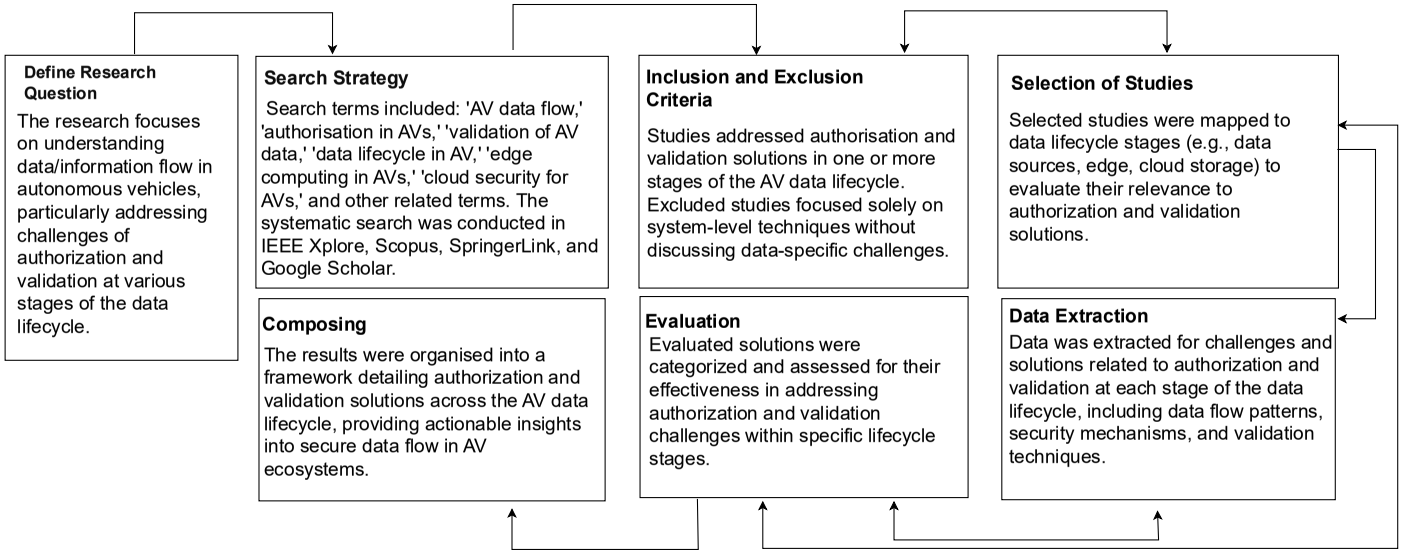

Fig. 1 Review Methodology. The flowchart outlines the key steps involved in conducting a systematic review of existing literature on AV. Double-headed arrows between some steps indicate bidirectional movement, highlighting the repetition of certain steps.

further enhance vehicle stability. The second generation of DAS, emerging in the early 1990s, sensors were classified as exteroceptive sensors (e.g., RADAR, LiDAR, and cameras) to detect the external driving environment [17]. The most recent generation of DAS, Advanced Driver Assist Systems (ADAS), has been developed to prevent collisions. A major turning point came in 2004 with the DARPA Grand Challenge, which accelerated AV research and development [18]. Subsequently, In 2006, the European Land-Robot Trials (ELROB) continued this movement by showcasing semi-autonomous vehicle capabilities. Major car manufacturers began developing automated vehicle technology in 2010. In 2014, Google introduced its first AV prototype [19], followed by Tesla integrated of automobile software for AVs in 2015. In the same year, and Ford commenced AV testing in California [20]. Alongside technological progress, regulatory frameworks have rapidly evolved. Between 2018 and 2020, the National Highway Traffic Safety Administration (NHTSA) proposed new AV guidance [21]. By 2021, Ford and General Motors had significantly invested in AV technology. Correspondingly, Robotaxi services were introduced by Chinese companies and tech giants like Baidu, Amazon, and Google in 2022 [22, 23]. The year 2023 saw the European Union developing AV regulations, the NHTSA issuing updated guidance for AV manufacturers, and China continuing to expand its regulations [24, 25]. In 2024, the UK introduced the AV Bill to establish safety regulations for AVs, setting the stage for these vehicles to operate on British roads by 2026 [26]. Additionally, In January 2025, the NHTSA proposed the AV Safety and Transparency Evaluation Program (AV STEP). This voluntary program invites vehicle manufacturers, Automated Driving System (ADS) developers, fleet operators, and system integrators to submit detailed information about their AV [27]. These developments reflect the global momentum toward integrating AVs into transportation systems. While an autonomous vehicle (AV) refers to the entire system, including body, mechanical controls, and user interfaces, the Automated Driving System (ADS) denotes explicitly the hardware and software responsible for performing dynamic driving tasks. For clarity, this paper refers to the ADS when discussing the technical system enabling vehicle autonomy.

2.2 AV Definition

Autonomous refers to a system's ability to change its behaviour in response to unanticipated events during operation [28]. According to NHTSA [29], an autonomous vehicle is one as which at least aspects of a safety-critical control function (e.g., steering, throttle, or braking) occur without direct driver input. However, vehicles that provide safety warnings to drivers (for example, forward crash signs) but do not perform a control function are not considered automated.

In September 2018, the NHTSA performed an extensive literature review of all the generic AV system features to identify the attributes that define the operational design domain (ODD). The comprehensive review resulted in 24 ADS features, specifically describing functionality and proposed timelines for commercial deployment across the different Society of Automotive Engineers (SAE) International levels of driving automation. Accordingly, the SAE's six-level taxonomy has become the widespread industry standard[30]:

Level 0: No Automation – The human driver performs all driving tasks, and any system support (like warning systems) does not automate driving.
Level 1: Driver Assistance – The vehicle may assist with a specific task, such as steering or acceleration, but the driver must remain in control and perform all remaining aspects of driving.
Level 2: Partial Driving Automation – The vehicle can control both steering and acceleration/deceleration, but the driver is responsible for monitoring the environment and must remain engaged at all times.
Level 3: Conditional Driving Automation – The vehicle can handle all aspects of driving in specific conditions or environments, but the driver must be prepared to take over when requested.
Level 4: High Driving Automation – The vehicle is capable of performing all driving tasks in specific conditions (such as certain road types or geofenced areas), and driver intervention is not required, though manual control is possible.
Level 5: Full Driving Automation – The vehicle performs all driving tasks in all conditions and environments, with no need for driver intervention at any time.

2.3 AV Architecture

Researchers studying AV focus on two main areas: defining their components and understanding their functional perspective. Some papers focus on the technical aspects of AV components, such as in [31], while others take a functional approach [32, 33]. From a technical perspective, AVs have two main layers: hardware and software. Each layer is comprised of several subcomponents. There is some disagreement among researchers about categorising the core competencies of different subsystems when defining the functional perspective of AVs.

However, in general, AV systems are made up of three to five primary functions: perception, localisation, planning, control and navigation, and system management [14, 34], as illustrated in Figure 3.

In the Figure, the several AV fundamental operations are as follows:

1. **Perception** refers to collecting data and extracting relevant understanding from the environment, such as the detection of road signs, as well as object detection and classification [35, 36]. These detection tasks are performed by various sensors such as cameras and Radio Detection And Ranging (RADAR).
2. **Localisation** refers to the ability of the AV system to determine the vehicle's position and orientation relative to the environment.
3. **Planning** consists of three stages:
   - Using algorithms, the path planner calculates the most efficient geometric path.
   - The behaviour planner determines the optimal behaviour based on the path planned by the path planner.
   - The estimation of the best possible route subject to vehicle dynamics and environmental constraints [32].
4. **Control** refers to executing planned actions and managing the vehicle's motions, such as changing lanes [14, 35].

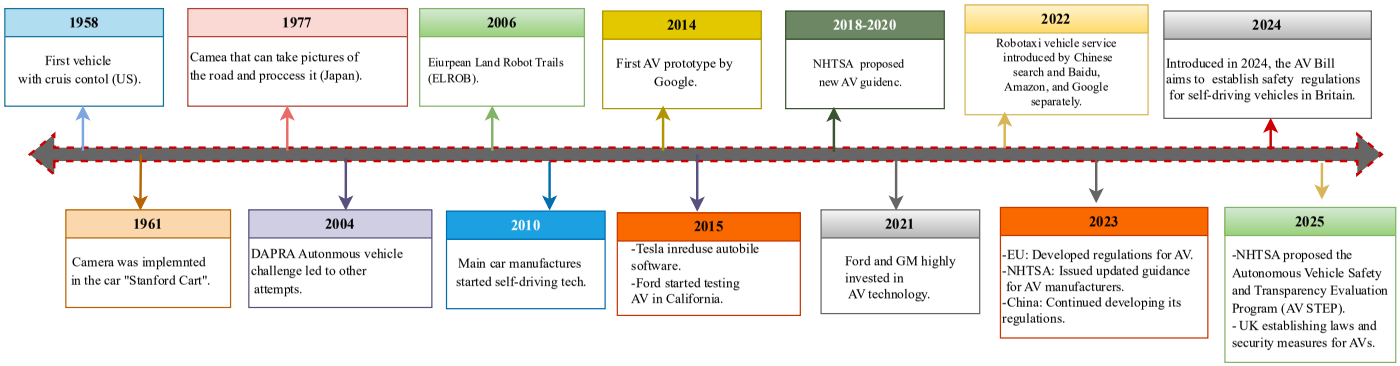

Fig. 2 The timeline of advancement of the autonomous vehicles in the last few decades. This timeline highlights the ongoing development of safety regulations and standards for AVs in various countries.

5. **System management** includes all the functions related to event data recorders, human-machine interactions involving in-vehicle interfaces [14], and external human-machine interfaces.

Fig. 3 Autonomous Vehicle (AV) Main Operations This diagram illustrates the main operations of an AV. It highlights five key areas: (1) Perception, where sensors such as cameras and RADAR collect data and detect objects; (2) Localisation, which determines the vehicle's position and orientation; (3) Planning, involving path, behaviour, and route planning; (4) Control, which executes the planned actions and manages the vehicle's motions; and (5) System Management, responsible for event recording and human-machine interfaces

An alternative architectural approach, such as in [37], divides AV systems into two primary components: the perception system and the decision-making system. While the perception system is further broken down into subsystems in charge of tasks like localisation, static obstacle mapping, and moving obstacle detection, the decision-making system is segmented into tasks like route planning, path planning, behaviour selection, and motion planning. This approach organises the autonomy system of the AV by highlighting discrete levels of perception and decision-making. Architectural paradigms like end-to-end and layered architecture further classify AV system designs[38]. The layered architecture divides the system into perception, including simultaneous localisation and mapping (SLAM), planning, and control layers, ensuring a modular and structured approach. While the end-to-end architecture processes raw sensor data directly to output control commands using deep learning techniques, providing reduced system complexity.

Sensors and Sensors Fusion

For an AV vehicle to have a high-quality and real-time understanding of its surrounding environment, it must quickly and accurately detect, comprehend, and track all objects. As a result, relying on a single source to generate all necessary data is impossible. Multiple sensors are equipped to provide both perceptual and location views of the environment, allowing the vehicle to make real-time decisions. Sensors are devices that translate detected objects or changes in the surrounding environment into quantitative measurements for processing [33].

Sensors Types

AV systems, while having some variations, all consist of two primary types of sensors based on their operational principle: Exteroceptive and Proprioceptive. Exteroceptive sensors are external state sensors that are utilised to perceive the environment, such as calculating the distance to objects or light intensity from the surroundings of the system. Examples of these sensors include cameras, RADAR, etcetera. Proprioceptive sensors, on the other hand, are internal state sensors that capture the dynamic state and measure the internal values of a dynamic system. Examples of this type of sensor include Global Positioning Systems (GPS), encoders, accelerometers, etc. [33, 39]. The most commonly used sensors for AV are listed in Table 2. In addition, Ahangar et al. [40] provided a detailed comparison of sensors and their individual challenges. As each sensor has different advantages and limitations in other aspects, integrating sensors is necessary to obtain an optimal perspective, and this operation is known as "sensor fusion".

Sensor Fusion

The advantage of this operation is to combine the data originating from different sources to complete each other's functions to provide an improved outcome in some specific criteria and data aspects for decision tasks [41]. The best example to explain the benefit of this method is to fuse RADAR sensors and camera images, where these sensors have different strengths and weaknesses. RADAR is not affected by the illumination of the environment, but it is not able to provide accurate data regarding an object's body. On the other hand, the Camera's image could provide these data, however, it may provide conflicting data under some illumination conditions [42]. The initial step in data fusion is sensor calibration, which means notifying the ADS regarding the sensors' position and orientation to collect and associate the data in space and time [43]. While the various sensors capture data in relation to the same object, the produced data could be combined to obtain different information to achieve higher-quality output. Other explicit examples with details and fusion techniques are discussed by Campbell et al. [42] and Fayyad et al. [39].

Table 2 Most Commonly Used Sensors. The table lists the most commonly used sensors in AVs, categorised into exteroceptive and proprioceptive sensors. Exteroceptive sensors gather information from the external environment, while proprioceptive sensors provide information about the vehicle's internal state.

Exteroceptive Sensors	
Sensor	**Function**
Light Detection and Ranging (LiDAR)	Uses light beams to provide a 360-degree distance between an object and the car.
Radio Detection and Ranging (RADAR)	Uses radio waves to determine the distance between an object and the car.
Camera	provides images of the environment to interpret data.
Ultrasonic	Used for short-distance object detection, such as parking.
Proprioceptive Sensors	
GPS (Global Positioning System)	Locates the vehicle.
IMU (Inertial Measurement Unit)	Measures acceleration and angular rate.
Encoders	Provide feedback signals used in speed and/or position control.
Accelerometers	Measure acceleration.

2.4 Communication in Autonomous Vehicles

AVs are able to communicate with any compatible systems, including other AVs, infrastructure, and pedestrians. This communication is known as Vehicle to Everything (V2X) technology, referring to how the vehicle communicates with everything. The network must be continuously fast, reliable, and secure to achieve an efficient cooperative environment with minimal delay (latency). Specifically, there are two main tendencies used all over the world: the wireless standard 802.11p or mobile networks, especially 5G [44–46]. In the foreseeable future, sixth-generation (6G) wireless systems will be crucial for V2X communications in AV [47, 48].

The differences between conventional and automated vehicles are apparent; however, the concept of connected vehicles represents a distinct phase in automotive technology. While the literature sometimes blurs the distinction between connected vehicles and automated vehicles (AVs), connected technology is a critical step toward achieving full automation. A connected vehicle is part of the Internet of Things (IoT), enabling data exchange, software updates, and communication with other vehicles and infrastructure. In these smart vehicles, all electronic control units (ECUs) and onboard units (OBUs) are interconnected through multiple digital buses, such as the Controller Area Network (CAN), Ethernet, FlexRay, Local Interconnect Network (LIN), Media Oriented Systems Transport (MOST) and Bluetooth [49, 50]. Such capabilities with sensing, communicating with the surroundings, and controlling the driving tasks represent the AV. The Connected and Automated Vehicle (CAV) is the vehicle that performs automated driving tasks and connectivity with other vehicles, road users, the road infrastructure, and the cloud [51].

While conventional cars constitute the vast majority on today's roads, forecasts suggest a notable rise in the presence of connected and autonomous vehicles in the coming years. Expectations indicate that the number of connected cars will reach 700 million by 2030, while the number of AVs will exceed 90 million [49]. Correspondingly, Zhang et al. [52] address the problem of optimally controlling CAVs under mixed traffic conditions where both CAVs and conventional vehicles are together on the

roads. Other studies [53, 54] subsequently aim to address the challenges and opportunities that arise and to leverage the capabilities of CAVs to enhance traffic flow at unsignalised intersections while ensuring safety in a mixed traffic environment where both conventional vehicles and connected and automated vehicles (CAVs) coexist on the roads.

2.4.1 Communication Technologies

Communication technologies play a critical role in enabling AV to interact with their environment, improving safety, efficiency, and driving experience. In this section, we explore several key communication technologies that are vital for vehicular networking, including Vehicular Ad-Hoc Networks (VANETs), and their associated communication types such as Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I), and Vehicle-to-Everything (V2X).

Vehicular Ad-Hoc Networks (VANETs):
A group of vehicles connected via a wireless network [55]. This network is a basic part of the Intelligent Transportation System (ITS) framework to assist various connections:

1. Vehicle-to-vehicle (V2V) communication that allows the vehicles to communicate with each other and share the necessary information, such as traffic jams. To establish V2V communication, vehicles should have an On Board Unit (OBU), Omnidirectional antennas, sensors and actuators, and a Global Positioning System (GPS) [56].
2. Vehicle-to-Infrastructure (V2I) communication enables the vehicles to interact with the roadside units RSUs which are fixed devices installed next to the road covering a dedicated area. These communications are mostly conducted by using wireless dedicated short-range communications (DSRC), which aims to provide active safety and convenience services [51].
3. Vehicle-To-Everything (V2X) Communication allows the vehicle to communicate with other entities using technologies such as Cellular V2X (C-V2X), including 5G and 6G.

VANETs networks utilise different communication technologies based on their transmission range, which can be categorised into short, medium, and long-range communication as shown in Figure 4.

1. **Short-range communication** includes Bluetooth, Ultra-Wideband (UWB), and ZigBee, which are wireless communication technologies used for vehicle-to-vehicle (V2V) communication in dense traffic environments due to their low power consumption and short operational range. Bluetooth, especially Bluetooth 5, the latest version, offers low-cost, low-energy communication up to 200 m, but faces interference and delays in dense settings. In contrast, UWB provides robust, energy-efficient communication with strong obstacle penetration and resistance to multipath fading, ideal for non-line-of-sight environments. Relative to other wireless protocols, ZigBee is simple, energy-efficient, and supports self-healing multi-node networks, though it may suffer interference on shared channels [57].

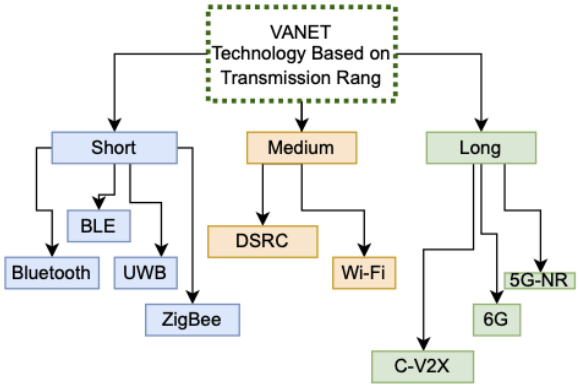

Fig. 4 VANET technologies based on transmission Range [40].

2. **Medium-range communication** relies on Dedicated Short-Range Communication (DSRC) and Wi-Fi, which provide higher data rates for V2V and V2I applications.
3. **Long-range communication** involves Cellular-V2X (C-V2X), 5G NR, and emerging 6G technologies, which support high-speed, low-latency communication for advanced AV applications.

Ahangar et al. presented a comprehensive survey regarding different vehicle communication technologies, their application, limitations, and advantages [40].
5G is suitable for long-range transmission to meet the high-mobility demand architecture; while 6G is still in its conceptualisation phase, although the technology-driven key performance indicates extremely serves the AV's communication requirements [58]. In general, various regions prefer DSRC over C-V2X (e.g., USA) due to the heavily deployed infrastructure around the country, however, this recently changed when Europe decided to move with cellular-based technology for CAVs [59].

In Brussels (2018), the 5G Infrastructure Public-Private Partnership (5G PPP), which is a cooperative combined initiative between the European Commission and the European ICT industry (ICT manufacturers, telecommunications operators, service providers, SMEs, and researcher Institutions, launched to deliver solutions, architectures, and standards for the next generation communication infrastructures [60]. The survey provided by Hakak et al. [61] highlights and summarises the key projects related to the 5G AV.

2.5 Autonomous Vehicle Potential Benefits and Costs

Automation technology is a promising future for safety, mobility, environment, and luxury [62]. It may reduce crash risks by avoiding human error and distracted driving. Many conventional vehicle crashes occur due to human error and distracted driving. In the USA, partially automated crash avoidance features could reduce the severity of as many as 1.3 million crashes every year, including 133,000 injury crashes and

10,100 fatal crashes [63]. Thus, traffic will disregard poor human driving behaviour and improve performance in terms of road safety and traffic congestion. In addition, AV contributes significantly to reducing emissions (worldwide goal to achieve net-zero by 2050 [64]).
Comparatively, even though adopting automation will increase safety, there are specific safety concerns related to its development. Its total or partial dependence on driving assistance systems results in a serious risk with both hardware and software issues. Additionally, sensors may be compromised due to environmental conditions such as dangerous weather. Another significant concern is that through cyber attacks, an automobile and/or its technological environment may be subject to causing grave privacy and security issues. Other benefits and costs are summarised in Table 3.

2.6 Data Authorisation and Validation in Autonomous Vehicles Ecosystem

Data plays a critical role in AV. Data is constantly being created, exchanged, and stored in this dynamic environment, creating challenges in data validation and authorisation that have not received the needed attention [65].

Table 3 Pros and Cons of applying Autonomous Vehicles.

Pros	**Cons**
Increased safety: Accidents will be greatly avoided due to the various assistance systems, ongoing connection, and connectivity between vehicles.	**Increased infrastructure costs:** AVs require higher standards for road maintenance and design.
Reduced energy consumption and pollution: Since these vehicles are supposed to run on sustainable energy, carbon and emissions of greenhouse gases will be almost nonexistent.	**System failure risks:** Hardware/software failures, wrong data, feeding/processing errors, faster traffic speeds, and increased overall vehicle travel are additional collision causes that may be on the rise.
Reduce traffic congestion: Although AVs move at a slower speed in cities, the traffic efficiency will be higher because of the efficient connection between vehicles.	**Data protection issues:** The network's environment causes security and privacy issues.

Data Authorisation

Data authorisation in the AV field refers to the set of mechanisms and policies that determine *who* can access specific data within an AV system, at *which* stage of the data

lifecycle, under *what* conditions and *how* this access occurs. These mechanisms mainly draw from well-established models such as attribute-based access control (ABAC) [66], role-based access control (RBAC) [67], and usage control (UCON) [68].

For instance, during a sensor maintenance operation, only authorised suppliers can access vehicle performance data for maintenance purposes in the required stage. Similarly, passengers' information, such as location history, may only be accessible to authorised parties (e.g., regulatory bodies) during an investigation, while remaining restricted at other stages to protect privacy.

Data Validation

Data validation in AV ensures that data used by the AV ecosystem at any stage of the data lifecycle is accurate, consistent, and reliable. This process includes sensor fusion, cross-verification, and cryptographic mechanisms to ensure trustworthiness [33, 69]. For instance, the system validates data from multiple sensors to confirm the detection of an object. Another example is validating accident-related data to ensure its accuracy and integrity.

Authorisation and Validation Requirements for AV Data

Based on inspiration from the literature, such as [70, 71], this work defines the Authorisation and Validation Requirements for AV as:

Requirement 1: Access permissions must be distributed among multiple stakeholder roles (e.g., manufacturer, supplier, user, regulator), each with clearly defined access rights.

Requirement 2: These access control rules and mechanisms should be adaptable, allowing for any changes in access rights, and customisable to ensure flexibility as the system grows and evolves.

Requirement 3: Stakeholders who generate data (e.g., passengers, owners, or manufacturers) must not have unrestricted access to all aspects of the data.

Requirement 4: Any critical access actions must involve approval from multiple independent stakeholders to ensure accountability.

Requirement 5: All these critical access actions must be recorded and traceable through secure logging mechanisms that are only accessible to authorised entities.

Requirement 6: Data validation processes must be conducted by domain experts, such as cybersecurity analysts and accident detectives to detect any unauthorised access attempts and ensure data integrity.

Requirement 7: Incorporated security measures must be applied as the data generated by external sources (e.g. V2X communication) may contain inaccuracies or malicious inputs.

Requirement 8: Data collected by the vehicle from the environment (e.g. pedestrians, other vehicles) may contain third-party information that the entity authorised to access the vehicle's data is not authorised to view or process. Mechanisms must ensure that such data is filtered or anonymised to respect the privacy and rights of external parties.

Requirement 9: Critical or personal AV data must not be leaked in any way to external, unauthorised entities (e.g. service providers or insurance companies) without explicit agreements.

3 Data Flow in Autonomous Vehicle Systems: Stages, Challenges, and Solutions

The operation of AV depends on a continuous and efficient flow of data. As illustrated in Figure 5, this data originates from various sources such as sensors and external environments, undergoes multiple layers of processing at the local and edge levels, and eventually is stored or analysed in backend systems. Each of these stages presents challenges that are unique, common, and intersecting and have been discussed in this section along with current solutions.

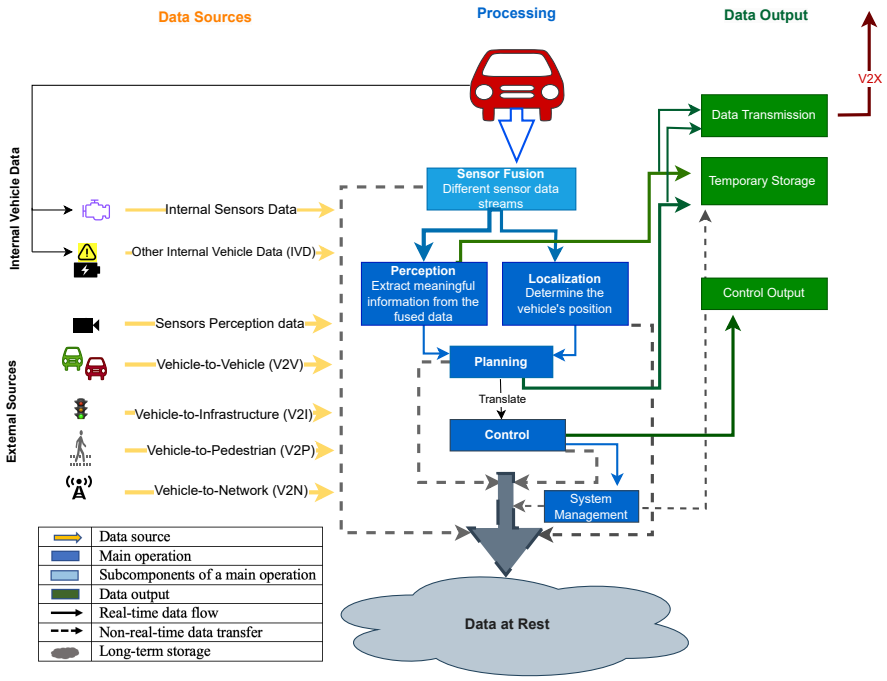

Fig. 5 Data Flow in an Autonomous Vehicle This diagram depicts the data flow in an AV system. It shows how sensor data and external sources are fused and processed through the Perception, Localisation, Planning, and Control stages. The outputs include Control commands, Data Transmission (V2X), and Temporary Storage, with all processes contributing to long-term Data at rest storage.

3.1 Data Sources

The vehicle's hardware and software systems collect and process substantial data to operate safely. These data sources might be categorised as sensory observation, Intelligent Transportation System (ITS), Geographic Information System (GIS) and map-related sources, social media feeds of drivers or passengers, linked data, archive, and legacy data [72]. As shown by yellow arrows in Figure 5, these data sources may alternatively be categorised as:

– **Internal Data Sources**, which are collected and processed within the vehicle itself, include:

1. Internal sensor data: These sensors monitor various parameters and parts of the vehicle's internal systems, such as engine and brake sensors.
2. Other internal vehicle data (IVD): This includes other data, such as the data about the owner and data that comes from the vehicle's status; for example, if the tank or battery is almost empty.

– **External Data Sources** are obtained from sources outside of the vehicle itself, which are crucial for various aspects of driving, navigation, and safety. In the same Figure 5 these sources include:

1. Vehicle-to-Vehicle (V2V) represents sharing real-time information about other cars' speed, direction, and braking status.
2. Vehicle-to-Pedestrian (V2P) communication includes exchanging information between vehicles and pedestrians or other detected road users, such as cyclists.
3. Vehicle-to-Infrastructure (V2I) refers to the exchanged information between vehicles and roadside infrastructure or traffic management systems.
4. Vehicle-to-Network (V2N) represents the connection between the vehicle and the cloud network to access navigation services, real-time traffic updates, weather, and entertainment content.
5. Vehicle-to-Everything (V2X) refers to the communications between vehicles and other entities, including cars (V2V), pedestrians (V2P), infrastructure (V2I), and vehicle-to-network (V2N). In the figure, V2X represents the comprehensive output integrating processed data from V2V, V2P, V2I, and potentially other sources, facilitating interaction with the broader environment, including vehicles, pedestrians, and infrastructure.

Key Challenges

1. Volume and Variety: The vast amount of data generated by AV poses significant challenges at every stage of its lifecycle. For instance, cameras produce 20–40 MB of data per second, while Light LiDAR systems generate between 10–70 MB per second [73]. The overall data volume and variety become extreme challenges with numerous heterogeneous devices and diverse communication streams. In addition to this wide variety of types and formats, the AVS must process in real-time to support effective driving decision-making. Another challenge is the authorisation to access and use data from these multiple heterogeneous devices and systems, especially when integrating third-party sensors or external communication streams.

Since these sources often rely on data from other devices, ensuring secure and controlled access and sharing is crucial to maintaining trust and safety in the AV ecosystem.

2. Data Quality and Reliability: Robust data validation is vital to maintain data integrity, especially in such environments with a high potential for noisy or incomplete data. In addition, the flow fusion information from multiple sources must identify and filter out corrupted or redundant data. Without adequate data validation and fusion, invalid or incorrect inputs could affect decision-making, potentially leading to unsafe driving behaviours.

Current Solution

Many recent studies have sought to ensure reliable data management, integrity, and efficient communication among sources at this stage. This includes data exchanged between vehicles, RSUs, stations, and data that reaches the sensors inside the vehicle. Some research suggests that blockchain technology could effectively address these challenges and meet data management requirements at this stage. For example, a blockchain model has been proposed in [74] that utilises the Hashgraph consensus algorithm to facilitate decentralised and secure data sharing among nodes. Each node disseminates information through a gossip protocol, enabling rapid consensus and verification of data integrity. Furthermore, instead of data encryption, Changvala et al. proposed a method to hide the integrity of LIDAR and RADAR data. [75]. Similarly, JAVED et al. [76] proposed a protocol to isolate false data in V2X communications messages. Another data integrity verification scheme is proposed in [77]; it aligns GPS data with other information regarding passengers to make sense of the vehicle's reliability. In addition, leveraging RSUs allows this data integration to be used for integrity checks. Another approach [78] focuses on a hybrid GNSS data compression method for autonomous vehicles, enhancing data transmission efficiency and resilience through frame differencing and entropy coding. Although its applicability may be limited, it achieves a good compression ratio and maintains reliability.

Additionally, in a notable attempt to improve the quality of sensors' data, Min et al. [79] have proposed a framework with two methods: first, a residual consistency checking algorithm that utilises sensor redundancy to isolate faulty sensors, and second, a Denoising Shrinkage Autoencoder (DSAE) that enhances anomaly detection in sensor data. Despite the algorithm's inability to isolate the "Spike" anomalies due to their brief duration, these methods help ensure that the sensors in AV are working correctly and provide reliable data. Similarly, to ensure that only trustworthy data is used for decision-making in AV operations, a study has integrated data quality metrics with a trust and reputation model[80]. This mechanism evaluates the correctness and reliability of real-time data sources based on their past behavior and interactions. While these solutions improve authorisation and data validation for AV data sources, further research is still needed in these areas and to address challenges like interoperability in data resources.

Table 4 Sample Event Data Recorder (EDR) Parameters Captured During a Vehicle Event [81]

Data Event	**Value**
Maximum Delta-V. Longitudinal (km/h)	-61
Time To Maximum Delta-V. Longitudinal(ms)	95.0
Maximum Delta-V Lateral (km/h)	-1
Time To Maximum Delta-V Lateral (ms)	72.5
Time To Maximum Delta-V Resultant (ms)	95.0
Ignition Cycle At Event	271
Ignition Cycle Runtime (minutes)	10.3
Odometer At Event Time Zero (km)	30.5
Airbag Warning Lamp Status	Off
ABS Warning Indicator Status	Off
Vehicle Drive Mode	Natural
Driver Safety Belt Status	Buckled
Passenger Safety Belt Status	Buckled
Occupant Classification Status in Front Passenger Seat	Small Adult
Driver Seat Track Position	Rearward
2nd Row Left Safety Belt Status	Not Buckled
2nd Row Left Seat Occupant	Not Occupied
2nd Row Center Safety Belt Status	Not Buckled
2nd Row Center Seat Occupant	Not Occupied
2nd Row Right Safety Belt Status	Buckled
2nd Row Right Seat Occupant	Not Occupied
3rd Row Left Safety Belt Status	Not Available
3rd Row Left Seat Occupant	Not Available
3rd Row Right Safety Belt Status	Not Available
3rd Row Right Seat Occupant	Not Available
Driver Airbag Deployment 2nd Stage Disposal	Yes
Right Front Passenger Airbag Deployment 2nd Stage Disposal	Yes
Complete File Recorded	Yes

3.2 Local Data Storing

Based on sensors and cameras positioned in various regions inside and outside the automobile, the vehicle creates and maintains data. According to how many ignition cycles the car goes through, most of them will hold their data for a while before replacing it with fresh data. Local storage systems, such as sensor data buffering, communication cache, and artificial intelligence (AI) models of data, play a critical role in managing large volumes of data generated by AVs. However, this section focuses on the event local recorder tools. Besides the data sent to the backend servers and cloud, an Event Data Recorder (EDR) and Data Storage System for Automated Driving (DSSAD) are the primary tools for storing significant event data. The rest of this section focuses on these technologies, their functions, gaps, and current solutions.

3.2.1 Evolution of Transportation Data Recorders

The first practical transportation data recorder was introduced in 1921. It records vehicle speed, engine RPM, and distance moved onto a rotating circular chart [82].

Figure 6 shows the current development movement of the existing national and regional activities on this technology. The diagram outlines the global timeline of the

adoption of the EDR. While South Korea and Japan were early adopters, the EU mandated EDRs for new and registered vehicles, and the US has ongoing discussions about implementing them. International organisations like SAE and IWG are working on developing standards for data loggers, including EDRs

3.2.2 EDR

Event Data Recorder (EDR) is a recording device inside the vehicle that can capture information regarding an event [82] in a readily usable manner. Its concept is the same for both AV and conventional vehicles; there is a list of all vehicles with EDR in [86]. Such devices would support real crash investigations and analyses of performance due to their data capturing of a few seconds, both pre- and post-crash. Useful and meaningful data, such as vehicle speed and brake status, are stored. For example, the Tesla Model 3 (Autopilot) is designed to record data such as how various vehicle systems were operating, whether or not driver and passenger safety belts were fastened; how far (if at all) the driver was depressing the accelerator and/or brake pedal; and, how fast the vehicle was traveling [81]. Table 4 is an example of such recorded data in detail. The technique behind this device generally stores data with the airbag control module once deployed. Other recording techniques and event data recorders have been explained comprehensively in [82].

In December 1996, the National Highway Traffic Safety Administration (NHTSA) and the National Aeronautics and Space Administration (NASA) cooperated to improve airbag safety. Consequently, NASA recommended that NHTSA encourage installing and obtaining crash data for safety analyses from vehicle crash recorders [82]. It is worth indicating, however, that (EDR- AV), the EDR for automated vehicles, must considerably exceed the performance of EDR for conventional vehicles (EDR-CV) [87]. Publications worldwide seek to provide "guidelines for the development of automated driving functions", including identifying the technical requirements for EDR-AV. The Regulation (EU) 2019/2144 sets rules on technical requirements for new types of motor vehicles to maintain safety and environmental protection [88]. According to that regulation, in June 2022, the EDR is supposed to become mandatory for all vehicles with SAE Level 3/Level 4 functions [87, 89]. From July 2024, all newly registered passenger cars in the EU must be equipped with an EDR [89]. In addition, the IWG EDR/DSSAD deals with the definition of the technical requirements as a prerequisite for the corresponding UN regulations for both EDR systems (conventional/autonomous) [90].

All crash data must be stored, available, and retrieved for crash reconstruction purposes, even for testing, to understand the crash circumstances [91]. However, for AV level 3 and above, where the system can perform all driving tasks for a specific time, EDR is not helpful regarding responsibility and liability without the Data Storage System for Automated Driving (DSSAD) system to indicate who was in control during the accident.

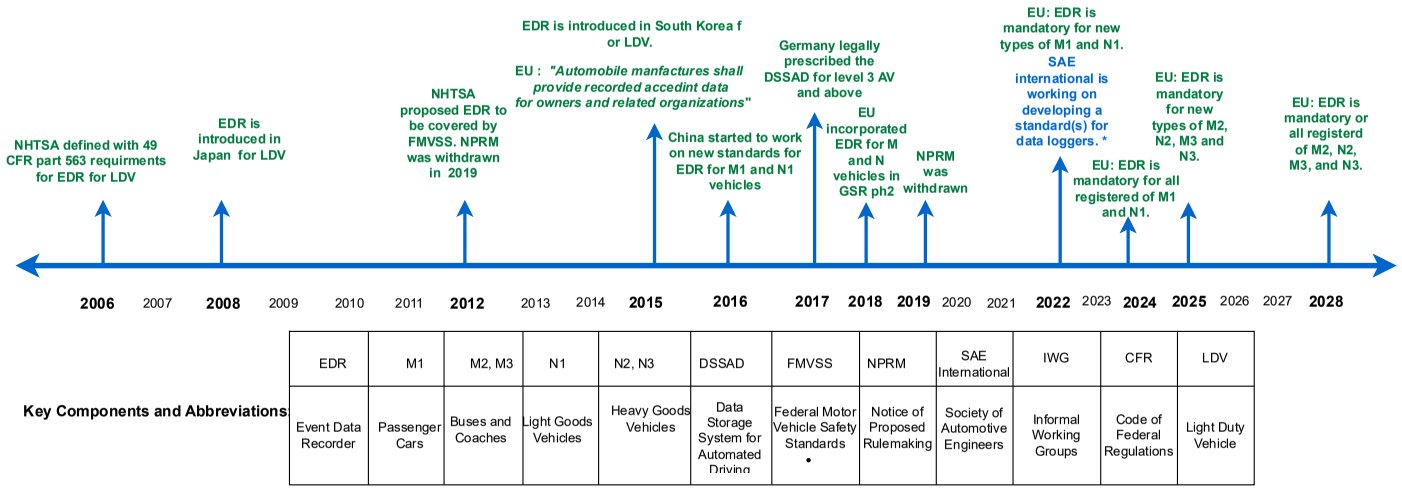

Fig. 6 Timeline of EDR and DSSAD Regulations and Standards Development (2006–2028) [83–85]. This figure illustrates key events in the development and implementation of regulations and standards for EDR and DSSAD across various regions.

3.2.3 DSSAD

Data Storage Systems for Automated Driving (DSSAD) is a storage device that determines whether the system or the driver has controlled the vehicle [84]. This provides information regarding who controlled the driving task at a particular time to decide the responsibility and, consequently, liability issues. Examples of stored data elements include position and time-stamped switches of the ADS from one mode to another [83]. These may contain data about whether the system is activated, manually or automatically deactivated. Therefore, it also records time and position and when the driver is requested to drive; position and time-stamped override through brake control by the driver; position and time-stamped transition demand by the ADS. Thus, the nature of the reasons for a transition demand or deactivation can be determined, such as:

- Driver not available or lack of driver attention.
- Driver override.
- System failure.
- Unplanned event (ex: bad weather).

While EDR is event-based storage, DSSAD is the continuous storage of specific AD data sets. All of these data must be clearly identified and recorded to eliminate confusion or misinterpretation.

3.2.4 Limitations of Traffic Accident Data Records

Despite the important roles that these tools play in supporting investigations, they have the following limitations:

1. Storage and Data Availability:
   Due to storage limitations and the large number of continuously variable parameters required for analysis, EDR stores only specific data parameters for a short time in specific events, such as events with airbag deployment. In addition, if these tools do not have sufficient space to record an event, they will overwrite or erase the previous event data.
2. Event Recognition Gaps: In particular, many accidents involving pedestrians and cyclists are not recognised as significant events, with no airbag deployed, meaning no event data is stored. Even when EDRs activate for crashes or near-crash events, they may fail to capture events that could be considered criminal but are classified as insignificant from a regulatory perspective.
3. Authorisation and Accessibility Issues: Ownership of EDR data is legally complex and varies by jurisdiction. For example, in the USA, ownership of the EDR data is a matter of State law, with some states considering the manufacturer to be the owner of the data, rather than the vehicle owner. While courts can obtain EDR data through court orders [91], NHTSA considers the vehicle's owner to be the owner of the data collected by EDR. However, the accessibility of this data introduces further issues. In some cases, companies such as Tesla and BMW have recently allowed the vehicle owner to access all EDR data using some unique hardware and software. Until 2017, they were the only parties permitted to access the data [81]. Furthermore, the original equipment manufacturer (OEM) may also access the

EDR remotely in specific crash scenarios, raising concerns about who truly controls the data.

These legal and accessibility challenges represent obstacles to access control. Although stakeholders, such as manufacturers and vehicle owners, may require access to the data, unrestricted access presents liability risks. The need for controlled access becomes apparent, allowing stakeholders to access the data under well-defined conditions and controls could be significantly valuable. In addition, DSSAD systems require consumers' approval for recording and/or accessing their data [83], highlighting the need for a balanced approach to access control. Therefore, legislators, regulators, and manufacturers play a pivotal role in determining what data should be recorded, how the AI could select the valuable data, the expansion of record device capacity, and, most crucially, who should be authorised to access it.

4. Insufficient Data for Forensic Analysis:

   Current tools cannot collect all five data classes typically required for forensic investigations: firmware, communication data, user data, safety-related data, and security-related data [87].

   A survey in which 173 international experts in accident analysis participated depicted that a considerable number of traffic accidents involving ADAS cannot be reconstructed [92]. Furthermore, the current EDR has to be refined to provide sufficient data for liability purposes [93] as the current data available is insufficient for thorough forensic analysis.
5. Privacy Concerns: Continuous recording and storage of video, location, speed, and/or surroundings of a vehicle appear to contradict the regulations addressing privacy protection. Therefore, laws related to this matter need to be enacted before it is adopted.
6. Need for Standardisation: Although manufacturers continue to develop new generations of EDR and DSSAD, international standards are urgently needed to ensure robust data validation and enhance their reliability. Global collaboration is needed to establish regulatory frameworks, standardise data elements, ensure interoperability, and define access authorisation.

3.2.5 Improving EDR and DSSAD for AV Systems

This section explores current efforts, recommendations and solutions for improving EDR and DSSAD for AV Systems.

Recommendations/Current Efforts

To address the limitations surrounding EDRs and DSSAD, various international efforts aim to improve current practices:

1. Time Window for Recording: The Aggregated Homologation proposal for Event data recorder for Automated Driving (AHEAD), which is a working group focused on developing standardized data models for recording information from a vehicle's Event Data Recorder (EDR) specifically designed for investigating accidents involving automated driving vehicles AHEAD, recommends data recording from 30

seconds before to 10 seconds after the collision. This time proves sufficient for the individual EDR-relevant claims of the present claims collective.
2. Inclusion of Vulnerable Road Users: Given the frequent involvement of pedestrians and cyclists in accidents, the Netherlands plans to immediately extend the scope of the EDR to include them.
3. Incremental Adoption of Standards: The UK government proposes a two-step method. First, introducing the USA's standards for EDRs in a relatively short timeframe; then, the second step applies more extensive requirements if adopting a single-step approach is not feasible. [94].
4. Various studies in the literature highlight critical requirements for local data storage in AV, such as those proposed by Kim et al. in their works [95, 96] and by Ten Holter et al. in [97]. This research and similar efforts will ultimately enhance safety and accountability in AV.

Technical Solutions:

This section presents various solutions for accident investigations and data integrity, as summarised in Table 5 that compiles solutions from multiple studies.

Various safety analysis models, including CAST [56] and FRAM [98], are used for accident investigation. CAST employs system theory to identify causes of failure and propose preventive measures. FRAM assesses complex interactions in socio-technical systems. These models lack a unified framework covering all causal factors, and manual analysis remains costly and inefficient.

According to [92], the solution to the limited information gathered independently from the manufacturer is a Forensic Event Data Recorder (FEDR). FEDR is an EDR that meets all the requirements from the investigator's perspective.

Data integrity preservation for investigation purposes in AV has been a goal of several studies that have presented frameworks based on various forms of technology. Hoque and Hasan have proposed a forensic investigation framework for AVs called AVGuard tool [99] that is designed for integration with the AD system. The framework assumes that the AV has local storage to store the log provenance while also communicating with a remote cloud server to publish the newly created log provenance. A robot operating system (ROS) node collects all the logs from different AD modules. Oham et al. have proposed a distributed digital forensics framework [100], which is based on the evidence reported by nearby witness vehicles if a vehicle is involved in an accident. Digital signatures, along with a corresponding certificate, are used to protect data integrity. Data exchanges between entities in the framework are stored in a blockchain and used for later decision-making. Further, T-Box [101] is a trusted real-time data recording system, which consists of an automotive data recording system with a network monitor, generator, and recorder. It assumes that the gateway could be used as a network monitor. The generator reconstructs data provided by the monitor and delivers it to the recorder, which stores data. The recorder stores an individual data entry into a block, and these data blocks can be stored locally or externally or transmitted to a remote server. Buquerin et al. [102] have provided a general concept for automotive forensics. Using Ethernet, their implementation uses the onboard

diagnostics interface, the diagnostics over internet protocol, as well as the unified diagnostic services for communication. Liu et al. aim to store EDR data safely and away from manipulation. In their scheme [103], data is not only sent to the manufacturer's server as usual, but the vehicle also uploads the EDR data to a cloud server and sends the evidence of storage to the nearby vehicle through a vehicular ad hoc network.
To conclude, while various solutions have been proposed to enhance accident investigations and ensure data integrity in AV systems, significant challenges remain. Issues such as data ownership, and privacy concerns still need to be addressed to develop a comprehensive and reliable approach for forensic investigations in AV environments. Future research should focus on integrating these solutions into a federated framework that balances accessibility, security, authorisation, integrity and regulatory compliance.

3.3 Local Data Processing

The collected data feeds the onboard diagnostics, which are part of the ADS function, i.e. driving decision-making, analysis of crashes, or technical failures. In addition to managing data collection from sensors and the fusion of sensor inputs, these embedded systems handle high-precision functions such as localisation, storage and updating of maps to finally perform complex tasks such as real-time control and machine learning. The data flow among these embedded systems faces many challenges, including integrating data across disparate AV systems, efficient communication [106], safety, and cybersecurity challenges [107]. These challenges require advanced solutions that optimise data processing efficiency.

3.4 Edge Processing

There are three cloud processing approaches: centralised location, edge-based processing, flexibility, and the Hybrid approach solution, which is the most preferred. Even though clouds are crucial for the success of AVs, there are several challenges facing the vehicular cloud community that are not faced by traditional cloud computing, such as high data transfer demands, latency concerns, and security risks. These factors drive the need for edge-based processing to offload computational tasks and reduce latency.

Key Challenges in Local and Edge Data Processing

Local and edge data processing plays a crucial role in ensuring low-latency real-time decision-making, but it also presents various key challenges that must be addressed to optimise data management, performance and reliability.

1. Latency-Sensitive Data Processing: While the low-rate data are regularly transmitted to the manufacturer's cloud, where all processing occurs at a centralised location, the high-latency-sensitive data, however, needs high-speed data processing within a distributed architecture.
2. Dynamic Resource Allocation: Unlike traditional cloud computing, vehicular clouds face challenges due to the dynamic nature of resources.
3. Lack of Central Authority: Managing security, privacy, authorisation, and authentication is more complex without a centralised authority [108], [109].

Table 5 Solutions for Accident Investigations and Data Integrity

Category	Solution	Description	Strengths	Limitations
Safety Analysis Models	CAST [56].	Uses system theory to identify failure causes and propose preventive measures.	Identifies system failures systematically.	No unified framework; costly and inefficient.
	FRAM [98].	Analyses complex interactions in socio-technical systems.	Handles complex interactions.	No unified framework; costly and inefficient.
Forensic Investigation Frameworks	AVGuard Tool [99].	Integrates with ADS, collects logs via ROS and publishes to a cloud.	Supports modular log collection.	Assumes reliable local storage.
	Distributed Digital Forensics Framework[100].	Relies on nearby AVs and Blockchain to ensure data integrity.	Uses Blockchain for tamper-proof evidence.	Requires witness AVs; may add network overhead.
	Automated Vehicle Data Pipeline. [104].	A pipeline consists of collecting raw sensor data and processing to reconstruct crash scenarios.	High-fidelity crash reconstruction.	Privacy and Security concerns.
Trusted Data Recording	T-Box [101].	Real-time data recording system with network monitoring and storage options.	Reliable real-time operation.	Lacks privacy considerations.
	Generalized Automotive Forensics [102].	Uses diagnostics and Ethernet-based communication.	Efficient diagnostic communication.	Limited implementation details.
Forensic Data Integrity	Safe EDR Storage [103].	Uploads EDR data to the cloud and shares evidence with nearby vehicles.	Prevents data manipulation.	Relies on vehicular ad hoc networks.
Forensic Data Integrity with Controlled Access	AVChain [105].	Blockchain and IPFS for secure, verifiable crash data sharing among stakeholders.	Ensures data integrity and controlled access.	The architecture's complexity and limited real-time capabilities
	Forensic Event Data Recorder (FEDR) [92].	Meets investigator requirements by gathering independent data.	Enhances forensic investigation.	Requires widespread implementation.

4. Network Bandwidth and Scalability: Large amounts of data require significant network bandwidth and scalable infrastructure for local processing[110].

Solutions for Local and Edge Data Processing

The edge-based cloud data centre performs many tasks at the edge instead of the cloud, leading to faster access than cloud computing. Edge and fog computing play a critical role here because the data requires high network bandwidth, which provides data processing and local storage capabilities. In addition, combining centralised cloud computing with edge-based solutions balances latency-sensitive and non-sensitive tasks.

An achievable paradigm for addressing issues with dynamic resource allocation and latency-sensitive data processing is mobile edge computing (MEC). MEC reduces reliance on centralised cloud infrastructures by bringing computational power closer to the data source, allowing quicker and more effective processing. This method works particularly effectively for ADS that must handle data securely and make decisions in real time. As summarised in Table 6, various studies propose solutions that optimise computation time, energy consumption, and resource usage while maximising privacy in edge environments.

Table 6 Mobile Edge Computing (MEC) Solution Studies – Summary of Objectives

Scheme	Minimising computation time	Minimising energy consumption	Optimising resource	Maximising privacy
[111]	✓	✓	✓	✗
[112]	✓	✗	✓	✗
[113]	✓	✗	✓	✗
[114]	✓	✗	✗	✗
[115]	✓	✗	✗	✗
[116]	✗	✗	✗	✓
[117]	✗	✓	✗	✗
[118]	✓	✗	✗	✓
[119]	✓	✗	✓	✗

3.5 Backend Cloud Computing

Most companies acknowledge and state that they keep a lot of data regarding the vehicle owner, the vehicle itself, and its in-vehicle hardware and software products [81]. The period of retention of this data varies from one company to another. They also differ in how they obtain this data; some transfer it physically to backend servers, while others do so remotely via networks. In addition, they differ in the data format; companies such as Tesla use raw sensor data, whereas some other companies ask Original Design Manufacturers (ODM) to provide processed data.

Captured and stored data requires substantial storage infrastructure, for instance, cloud or on-premises servers. The majority of automakers are utilising cloud-based capabilities via connected-car services. For example, in 2019, Ford publicised its connected-vehicle collaboration with AWS. In addition, Toyota introduced its engineering ecosystem in 2020 to develop and deploy the next generation of cloud-connected vehicles alongside similar initiatives outlined in [120].

Indeed, in a promising attempt to address the data management challenges in AVs, Cloud technology is a scalable solution in the automotive industry. Therefore, data is classified as onboard data, sent to the cloud or stored in servers and hard drives as long-term storage. Local function data and V2X data that has a 4ms response time requirement and has to go off-vehicle will not be sent to the cloud [121]. In contrast, model training data, for example, includes cases where a new object has been detected, and the data about this anomaly will be used to formulate patterns and reports will be sent to the cloud for future algorithm improvements [122], [123].

3.5.1 Key Challenges and Current Solutions

The following key challenges in backend cloud computing for AVs appear insufficiently addressed in the literature.

1. The Need for Data Sharing

Even though some vehicle manufacturers agree to share in-vehicle data with the other service providers by accessing data directly through the vehicle manufacturer's server or via "neutral" servers that would gather the data, the service providers ask for direct real-time access to in-vehicle-produced data and functions through an in-vehicle interoperable, standardised, secure, and open-access platform [124]. In addition to legal authorities, Tesla states that the data could be shared with their service providers, business partners, and affiliates; in addition to any third parties, the owner has also been authorised [81]. Some stakeholders suggest that car manufacturers should allow tier 1 suppliers to access the data directly [125] to maintain and improve their products. Figure 7 illustrates the tier levels within a manufacturer.

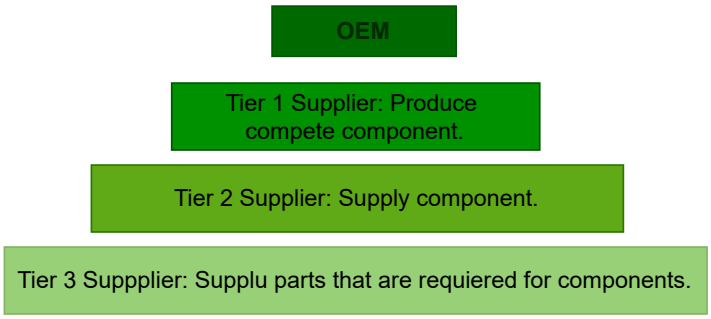

Fig. 7 Hierarchy of supplier tiers in automotive manufacturing.

Various situations are defined in the literature where stakeholders need to access accurate AV data, such as accident and failure data. In traditional vehicles, human error is the critical cause of 94% of vehicle accidents [126]. However, in the case of an accident with an AV, there are clear differences in liability. In particular, the interaction of many factors in AVs and their ability to make some or all driving decisions disrupts the adjudication process. The liability subjects of AV accidents may include vehicles, vehicle assistant drivers, manufacturers, vehicle owners, and insurance companies [127]. However, the need for this data from stakeholders encompasses a much broader scope than these parties. In the literature, authors and organisations differed in their identification of those stakeholders and the techniques they use to identify them. NHTSA states that motor vehicle accidents may be investigated through various entities, and they listed them [82]. Another example, the authors in [14] fragmented stakeholders into three general classes: A) containing all types of end-users and society; B) containing all technical groups; and C) for all regulatory parties, including insurers. However, the most frequent entities in literature are as follows:

1. Government and the legal authority: The accident data collected from the involved AV are necessary as facts and evidence in criminal courts to determine liability. Another possible scenario for government use is to take advantage of data to reduce the possibility of accidents in the future and thus reduce losses and assess roadside safety.
2. Original Equipment Manufacturer (OEM): The Manufacturers may analyse the collected data and use it to monitor system performance and improve their products. Another possible scenario is to reconstruct the accident to assess the causes in order to reduce the possibility of accidents in the future.
3. Suppliers: The potential scenario is to analyse and improve the products or services that are involved in a vehicle manufactured by the OEM.
4. The owner: In one possible scenario, the owner might use the data as evidence to absolve him/her of any criminal liability. In addition, the provided data could be used for insurance amounts and services.
5. Insurance: The stored data presents accurate evidence to resolve insurance disputes and a fair solution that ensures that no party is tampered with.
6. Testing and certification bodies: The testing organisation needs the data to reconstruct the accident in order to analyse and update the technical and legal requirements.
7. Road authorities: Data can support assessing and improving the infrastructure and roadside safety.
8. Researchers: To analyse and assess the AV crashes to assist in the development of vehicles, infrastructure, and the whole environment's components.

   In addition, based on this study's findings, a further entity has been identified:
9. Other Manufacturers: This category includes companies involved in manufacturing the vehicles. These manufacturers can leverage the collected data to improve their products, assess compatibility with AV technologies, and enhance training for AVs, especially when they encounter new cities or environments they have never experienced before. This strengthens the AV's adaptability and operational efficiency, contributing to the development of safer and more effective ADS.

Due to the significance of preserving the data generated and received by the system, it is clear that sharing this data with all relevant parties is essential.

Current Solutions

One approach that facilitates collaborative data sharing and mitigates privacy concerns is using Federated Learning (FL), which is a distributed collaborative AI approach that allows multiple devices to coordinate data training with a central server without sharing actual datasets [128]. With a trusted server (aggregator), parties can learn a shared machine learning model locally and separately, as well as share only the resulting insights from each analysis. This technology is an active area adopted recently in various applications and research such as in financial applications [129], mobile applications [130], biomedical research [131], which relies on Multiparty Homomorphic Encryption (MHE) to perform privacy-preserving FL by using the advantages of both interactive protocols and homomorphic encryption (HE). Another example is to encrypt only the critical parts of model parameters to reduce local computation

and communication costs. Sotthiwat et al. [132] proposed a partially encrypted Multi-Party Computation (MPC) solution that only encrypts the first layer of local models with MPC strategy.

In the AV domain, blockchain-based FL systems have been recently introduced to enhance security, transparency, and reliability in data sharing. For example, a distributed on-vehicle machine learning model [133] has been proposed to improve vehicular networks. Despite the increased computational overhead in that work, data could be trained, and models could be exchanged in a distributed manner. Similarly, a framework has been presented to enable vehicles to encrypt portions of their data using HE before uploading it to a cloud server[134]. Zeng et al [135] also designed an automated controller to avoid performance deficiencies of traditional learning-based controllers that are trained by each connected vehicle's local data. In their design, the learning models used by the controllers are collaboratively trained among a group of CAVs.

While these approaches and similar ones demonstrate potential solutions, they require further refinement to enable AV manufacturers to collaboratively benefit from FL algorithms by sharing knowledge across their respective clouds without exchanging any raw data. Moreover, they could be effectively integrated to facilitate collective training that would otherwise be unattainable individually.

On the other hand, while these algorithms preserve privacy, this advantage comes at the expense of the model's accuracy due to encryption and the limited operation set they support. In other words, approaches such as FL, HE, and MPC offer significant advantages in preserving data privacy and security, and play a crucial role in facilitating secure data analysis. However, in certain use cases, sharing partial or complete datasets is a desirable goal to enhance model accuracy and performance, allowing for more comprehensive insights and better decision-making while protecting sensitive information.

ADS in AV as an emerging and future-oriented field requires broader and more effective collaboration and data sharing among the stakeholders mentioned in the previous section. Thus, there is an urgent need for systems that integrate highly secure data sharing for AV. The literature presents several attempts to propose secure access control mechanisms in various data fields, such as [136], [137], [138]. Even though there are a few notable systems that have been developed and tailored to AV data [105, 139], similar works in this field are rare and nearly nonexistent.

2. Data Validation

The challenges of data sharing are not limited to direct access; rather, there is also a need to validate data. However, due to the large volume of heterogeneous data, it is quite difficult to even pre-process it effectively. Therefore, data validation becomes a critical step in this stage of the data lifecycle to ensure consistency, accuracy, and reliability.

Current Solutions

A blockchain-based platform with MPC is proposed for the AV data validation process [140]. BELIEVE, which stands for Blockchain-Enabled Location Identification and

Efficient Validation with Encryption approach, integrates real-time data sharing for immediate decision-making with backend storage on a distributed ledger to ensure that validated data is securely recorded and accessible for future reference. AV systems rely heavily on big data analytics, thus, data quality improvement strategies are critical to address challenges relevant to a big data environment, such as the management, storage, cleaning, integration, and reducing inconsistencies and optimisation. Data quality improvement studies have been proposed recently, such as [141–143]. These studies and further programmes are needed as essential parts of validating data quality before and after it is stored.

4 Cross-Cutting Concerns

The ADS operate in a complex and dynamic environment, generating vast amounts of data that must be processed, stored, and shared seamlessly. This section explores certain overarching concerns that impact the system as a whole. Addressing these concerns is critical to ensure reliability, trustworthiness, and ethical operation.

4.1 The Complexity of Fault Detection

An ADS generates a vast amount of data, making the extraction of crucial information from the logs of different AD modules a real challenge. In the case of a defect, finding the set of influencing factors causing the failure is a complex mission due to two reasons: first, the AD functions may not be sufficient for all unexpected conditions in the dynamic environment with unlimited contexts; second, deviation from the intended functionality due to the inductive nature of a system that combines machine learning components. Failure in sensor fusion, sensor readings (e.g. misdetection), external environment context (e.g.: weather), or control issues (for example: braking is not initiated in time), can result in unintended outcomes, sometimes due to combinations of these factors [144, 145]. Another proposed taxonomy by Zhao et al. [146] for potential sources of sensor data anomalies in AV is categorised into four main categories: faults in components, adaptability failures, cyber-attacks, and design deficiencies. Within the enormous amount of produced data, locating meaningful data that pinpoints the crash cause or causes is a significant challenge. Identifying the precise moment of failure, or the faulty device or subsystem, adds further complexity. In discussing the complexity of fault detection, various external factors contribute to scenarios that may lead to dangerous situations. Identifying these causes of harm complicates fault detection because it requires a broader understanding of the traffic context, not just the internal functioning of the vehicle.

Figure **??** illustrates the complexity of sources of harm in the ADS based on ISO 26262, ISO/PAS 21448, and SAE J3016 standards. It categorises harm arising from malfunctioning behaviours, functional insufficiencies, and deviations from the Operational Design Domain (ODD), the specific conditions where the ADS is designed to operate safely. In addition, it highlights that even with proper design intent, failures can occur, and the operational context influences these sources of harm. Furthermore, since crashes are rare events, testing and analysing similar scenarios to define the exact

harm that may be blamed on one party or another in this nested system is particularly challenging.

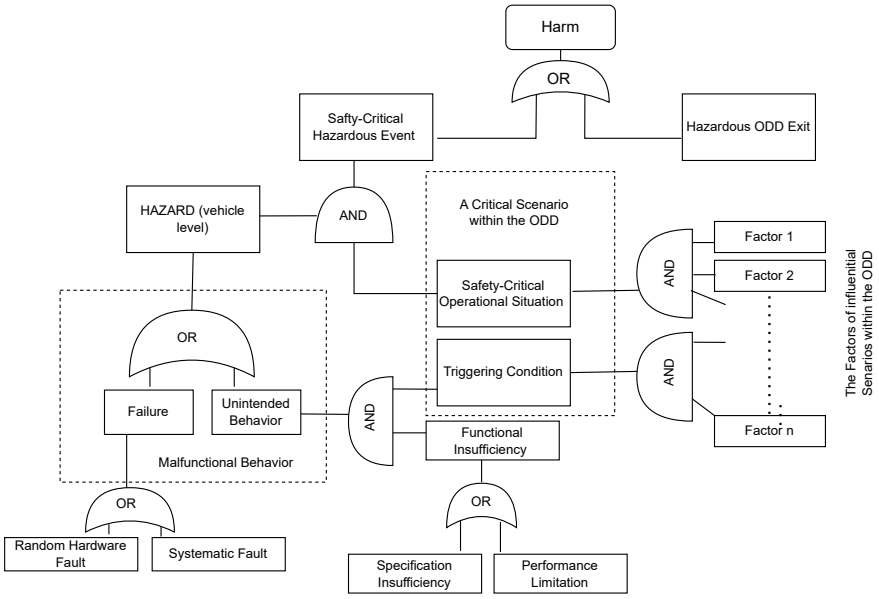

Fig. 8 The Complexity of Finding The Source of Harm[144].

For real-time safety assessment, [147] has proposed the Bayesian Hierarchical Spatial Random Parameter Extreme Value Model (BHSRP) for real-time safety assessment. This model could address the difficulty of extracting meaningful information from the massive and complex datasets generated by ADS.

In this context, a conceptual fault-handling system design for driverless trucks was proposed in [148], which highlights the need for advanced fault detection mechanisms in driverless vehicles. The findings in the paper emphasise the difficulty in determining the exact cause of a failure. Similarly, Koopman and Wagner [149] discuss the main challenge in testing AV, framing their analysis within the V-model developed under ISO 26262. This framework links various types of testing to ensure that safety-critical systems meet required standards. However, ISO 26262:2011, which sets standards for safety-critical systems, requires adaptation to accommodate the complexities of future autonomous driving technologies [150]. In the same context, A criticality analysis framework [145] aimed to identify and analyse critical traffic phenomena for the verification and validation of automated vehicles. The proposed mechanism involves a combination of both expert-based and data-driven approaches to identify relevant criticality phenomena and explain the underlying causes. These studies and similar ones lack a comprehensive approach to identifying influencing factors in such a complex and dynamic traffic environment. They also, as mentioned earlier, face limitations in model constraints and data availability when assessing critical situations.

Maintaining the high-quality training data sets is crucial for AI systems that support decision-making in AV [151]. Predictive models that leverage machine learning approaches can further enhance data validation by classifying useful and non-useful data, thereby improving real-time analytics. For example, the machine learning-based approach for predictive analytics, as in [152], offers possible solutions for identifying system failures, optimising data validation and addressing sensor misdetection and control failures in such dynamic environments. Further research is needed to address the challenges of identifying the source of harm and fault in AV due to the complexity of their systems and operating environments.

4.2 Data Privacy

The open wireless access in VANET seriously impacts not only the privacy of users, but also the pedestrians whose photos and locations can be captured by vehicles. AVs generate a vast amount of data that is collected, stored, transferred, and shared, posing severe risks to user privacy.

Privacy concerns span all phases of data flow in AVs and require thorough investigation. Table 7 summarises key studies addressing these challenges, detailing their aims, mechanisms, results, and limitations in providing privacy solutions for AV.

4.3 Security

Security in AV is a critical concern as network vulnerabilities can lead to severe consequences, including loss of life. Due to the complexity of AV systems, security threats target various components, and these attacks can be broadly classified into the following categories:

In-vehicle network

Attacks on In-vehicle networks include the following:

1. **ECU engine control units:** Attackers may compromise the ECU by altering programming code, affecting vehicle performance.
2. **CAN and SEA J1939 buses:** The CAN bus, which connects all the vehicle's components, is a critical target for attacks. For example, malicious actors can inject viruses into the CAN bus, disrupting critical operations
3. **Remote sensors:** Any tampering with the sensors' data generated and transmitted can result in fatal accidents. Through existing wireless networks, external entities can make connections with sensors, implementing remote sensor control.
4. **GPS:** Adversaries can alter GPS data, disrupting navigation and decision-making processes.
5. **Wireless communication:** Wireless technologies such as Bluetooth, tire pressure monitoring systems (TPMS), and keyless entry and ignition systems present additional vulnerabilities. Attackers may exploit these systems to gain unauthorised access or control over the vehicle [65].

Table 7 Summary of Privacy Concerns and Solutions for Autonomous Vehicles in the Literature

Paper	Aim	Mechanism	Results	Limitations
[153]	To preserve privacy in AV using homomorphic encryption during the storage and processing stages.	A Pixel-level encryption for secure searching over encrypted images with probabilistic trapdoors.	Reduces storage and increases efficiency.	Lacks detailed discussion on real-world implementation and performance impacts on various cloud environments.
[154]	To develop a privacy-preserving authentication scheme for V2G networks.	The scheme uses randomly selected pseudonyms for AVs and establishes a secure session key through an authentication key agreement protocol to ensure confidentiality.	The proposed scheme achieves lower communication overhead (800 bits) compared to existing schemes.	The paper does not provide real-world testing of the proposed scheme.
[155]	Enhance privacy in federated learning for AVs	Gradient encryption in federated learning to preserve user privacy without extra computational cost.	Improved accuracy (2% higher) and reduced data transfer compared to conventional FL.	Additional computational infrastructure needed for blockchain integration.
[156]	Propose a context-aware privacy-preserving method for AVs.	Uses SDN and differential privacy/data aggregation depending on data sensitivity.	Shows higher performance in privacy preservation, cost, and latency compared to existing methods.	May require further evaluation with third-party providers; potential limitations in highly dynamic environments; computational complexity and overhead considerations.
[157]	Develop a framework that allows users to choose which parts of their data they want to keep private before sharing it with other vehicles.	PRECISE framework utilises secure segmentation to identify sensitive objects, inpainting techniques to remove them, and edge computing to process data using secure deep learning models enhanced by additive secret sharing.	The framework achieved secure segmentation in 3.47 seconds and inpainting in 0.99 seconds.	Processing times may impact real-time performance, and the security of edge servers poses a risk of data exposure.
[158]	To improve traffic efficiency and fuel economy while protecting the privacy of vehicle data using cloud-based collaboration.	an affine masking-based privacy strategy, which encrypts vehicle state data before sending it to the cloud and decrypts the control input using inverse affine masking to protect privacy during vehicle-cloud collaboration.	The scheme enhances traffic efficiency and fuel economy while securely masking vehicle data through an affine masking technique	The limitations include potential computational overhead affecting real-time performance and scalability challenges when handling a large number of vehicles simultaneously.

6. **Denial-of-Service (DDoS):** Distributed Denial-of-Service (DDoS) attacks have emerged as a critical concern within the V2X and VANET environments. These

attacks flood vehicle communication channels with malicious traffic, leading to service disruptions, degraded safety performance, or even system failure.

Vehicle to everything network (V2X)

Attacks on V2X communications typically target the following systems:

1. **VALNET or Vehicle ad-hoc networks:** VANETs rely on dedicated short-range communications (DSRC) and are based on the IEEE 802.11p standard for wireless access in vehicular environments.
2. **Mobile Cellular Network, Satellite Radio, and Bluetooth:** Another communication structure required for V2X are the mobile cellular network, satellite radio, and Bluetooth, which can be targeted by attackers to disrupt vehicle communication or gain unauthorised access to vehicle systems.

Another taxonomy of attacks has been provided by Gupta et al. [159] based on the architecture of hardware, network, and software as in Figure 9.

Several solutions in the vehicular network have been recently proposed. For instance, RTED-SD, which is a scheme that aims to detect real-time attacks [160]. In this scheme, the authors use the Fast Quartile Deviation Check algorithm (FQDC) to recognise and locate the attack in the Internet of Vehicles. Similarly, a threat prevention framework has been proposed in [161] for Vehicle-to-Vehicle communication in AV Networks, integrating dynamic risk assessment using the Probability-Impact-Exposure-Recovery (PIER) metrics, security decay assessment via ruin theory, and a risk-aware message forwarding algorithm based on game theory. This approach aims to enhance security and privacy by proactively addressing vulnerabilities in V2V communication. Furthermore, to detect malware attacks in AV, Aurangzeb et al. [162] proposed a hybrid approach that combines static and dynamic analysis for real-time detection. The mechanism demonstrated malware detection that also enhanced safety and reduced communication latency. Additionally, a method proposed by Cretu et al. [163] is a method that utilises evolutionary search and machine learning to identify vulnerabilities in query-based systems (QBS). It showed higher performance in finding vulnerabilities compared to existing attacks.

Various works have explored and enhanced Intrusion Detection Systems (IDS) in AV, aiming to address different challenges and approaches. Anthony et al. [164] propose a method called NTB-MTH-IDS, a Non-Tree-Based Multi-Threshold Hybrid Intrusion Detection System, which is an intrusion detection system that leverages non-tree-based machine learning techniques. Similarly, Anbalagan et al. [165] introduce IIDS, which is based on a deep Convolutional Neural Network (CNN) system that transforms vehicular network traffic data into images for attack detection. Additionally, Aloraini et al. [166] investigate adversarial attacks on IDSs in in-vehicle networks. For detecting DDoS attacks in VANET, recent studies have investigated the machine learning-based techniques, for example, the approach presented by Setia et al. [167].

Another security approach for AV that has emerged as a promising solution is quantum encryption, which aims to face the potential threats posed by quantum computing. Despite the complexity of quantum-based security solutions, several

works have explored their application in AV. For instance, including blockchain-based authentication, quantum key distribution, and secure federated learning [168–170].

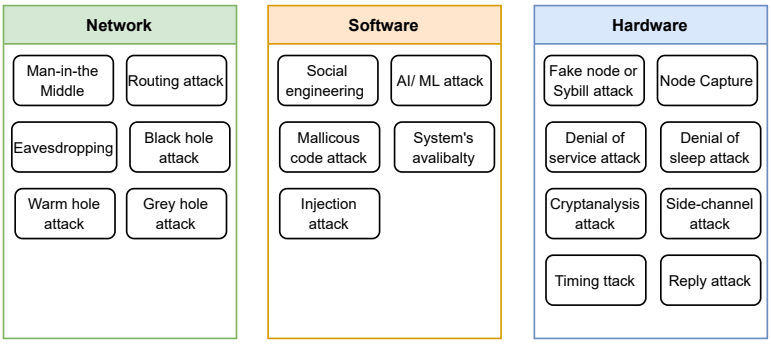

Fig. 9 Attack Taxonomy [159].

Various methods recently have proposed the integration of BC technology in IoV systems, aiming to enhance security. For example, [171] uses a consensus mechanism to provide a secure event-sharing protocol for smart cities using IoV and RSU. Similarly, [172] aims to protect IoV against quantum attacks. In addition, a blockchain-based mechanism [173] is designed to provide security-related data collection and incentivise mobile nodes.

Despite the advancements of these solutions, they still face notable limitations. In addition to their computational overhead and integration complexity, many approaches rely on predefined threat models, limiting their ability to detect unknown attacks. An overview of the key security solutions proposed for autonomous vehicles, including their applications, approaches, and limitations, is presented in Table 8. A thorough understanding of both vulnerabilities and potential solutions is essential to advance the security and privacy of AV in a connected landscape.

4.4 Regulation and Standards

Most of the current regulations that relate to human drivers must be changed to properly address AV, including testing and deployment, safety standards, data exchange standards, security standards, liabilities, insurance, and personal information privacy standards. For example, much of the current privacy legislation is inappropriate for AVs, such as the U.S federal Drivers' Privacy Protection Act, and Electronic Communications Privacy Act [94].

Indeed, some countries have started acting on new regulations; for instance, the Scottish Law Commission's regulation for AV and AV legislation and policies in the USA, the Netherlands, the UK, and Sweden [14],[174]. In addition, EU countries, industry, and the Commission are collaborating to achieve the EU's ambitious vision for connected and automated mobility across the EU. The commission will address many

Table 8 Summary of Security Solutions for Autonomous Vehicles in the Literature

Security Solution	Application Area	Mechanism	Results / Contribution	Known Limitations
RTED-SD (Real-Time Event Detection-Stop Detection)	Road event detection, Driver behaviour analysis	Hypergraph-based technique for analysing multi-vehicle interactions to detect risky events in real time	Enables fast detection of hazardous driving scenarios	High computational cost due to complex multi-vehicle data
Hybrid Malware Detection	Malware detection in V2X communications	Combining static and dynamic malware analysis in real-time systems	Enhancing malware detection while maintaining safety and reducing communication latency	Processing overhead and real-time complexity
Intrusion Detection Systems (IDS)	Detection of DDoS, adversarial, and spoofing attacks in vehicular networks	ML-based IDS (e.g., NTB-MTH-IDS), CNN-based traffic analysis, GAN-based adversarial attack detection	Improves detection of varied cyber threats using adaptive learning	High dependence on training data; costly model updates; computational intensity
Quantum Encryption	Secure communication and confidentiality	Uses quantum key distribution and federated learning for AV communications	Offers future-proof security resistant to quantum threats	High implementation complexity and cost of quantum hardware
Blockchain-based Authentication	Data integrity, privacy, and access control in V2X	Uses smart contracts and distributed ledgers for secure authorisation and tamper-proof logging	Enhances trust and transparency in AV systems	Scalability issues and computational overhead in high-speed networks
Blockchain-based Incentive Mechanism	Secure data sharing and collaboration in IoV environments	Incentivises honest behaviour in AV networks through token-based mechanisms linked to data integrity	Encourages cooperation while ensuring verifiability of shared data	Assumes predefined threat models; integration complexity

current issues, such as policies and legislation relating to digital technology, including cybersecurity, liability, data use, privacy, and radio spectrum/connectivity, which are of increasing relevance to the transport sector [175]. This supports the aforementioned 5G PPP; however, there are no international central regulatory bodies existing to regulate the implementation and deployment of AV. In light of these regulatory challenges, various global principles can offer valuable insights into shaping the future of AV regulation, such as the Precautionary Principle (PP), the Principle of Preventive Action, and the Best Available Techniques (BAT) Principle. There is a growing need for greater academic investigation into the best legal and regulatory practices for ADS.

Recent studies [176–178] have suggested frameworks and approaches to guide the development and regulation of AVs and their related technologies. Similar and further efforts are needed to help balance the benefits and risks of AV by incorporating safety and ethical considerations into policies and standards.

4.4.1 Cross-Border Interoperability

Accessing diverse services and sharing data among vehicles and infrastructure is a critical component of vehicle decision-making. However, interconnectivity and interoperability remain significant challenges in our geopolitically partitioned World.

Cooperative Intelligent Transportation Systems (C-ITS), a paradigm that is based on Information and Communication Technologies (ICT), enables the creation of both stand-alone in-vehicle systems and cooperative systems (V2X) [179]. While C-ITS solutions provide valuable services, the deployment of their infrastructure and delivery of their services often encounter territorial and regulatory hurdles [180]. An architecture has been presented in [181] that addresses these challenges by focusing on the integration of cooperative intelligent transportation systems in automated driving with an emphasis on cross-border interoperability. This AUTOCITS architecture was implemented in three European cities, showing the potential for harmonized systems across national boundaries. Achieving seamless cross-border interoperability in the ADS of AV requires significant collaboration among stakeholders, including governments, industry, and international standardization bodies. Bridging regulatory and technical gaps across regions remains essential for realising the full potential of C-ITS and AV ecosystems.

5 Conclusion and Future work

This paper highlights the essential role of data authorisation and validation in Autonomous Vehicle ecosystems, reviewing recent advances at each stage of the data lifecycle. As AV technology evolves, traditional data frameworks must adapt, with technologies like blockchain, zero-knowledge proofs, and federated learning playing increasingly vital roles. Global regulatory collaboration and integrated systems are also essential to ensure the safety, privacy, and security of AVs and their connected infrastructure. The key findings and insights of this review are as follows:

1. Data Flow, Integrity, and Accessibility: Secure data management is fundamental for ADS. The flow of data, from acquisition and processing to sharing and storage, must be protected from unauthorised access and manipulation to ensure system reliability and safety. The integrity of this data is critical, not only for real-time decision-making but also for accident reconstruction, forensic investigation, liability, production development, research, regulatory, and for all parties involved in this operation cycle. Modern technologies such as Blockchain can provide immutable records for data validation, ensure traceability in data exchanges, and enhance transparency in AV systems at each stage. However, in most frameworks, it does not inherently provide access control or authorisation, which must be handled by other mechanisms.
2. Data Ownership, Ethical Considerations, and Authorisation: The data ownership in AVs is complex, with multiple stakeholders, including manufacturers, governments, and users, each having a vested interest in the data generated by these systems. This complexity presents obstacles in sharing solutions due to ethical considerations and conflicting interests. Therefore, any data authorisation process must ensure that only authorised entities can access or modify critical vehicle data to strike a balance between innovation and privacy concerns. Multi-factor authentication (MFA) strengthens identity verification, zero-knowledge proofs (ZKPs) allow verification without exposing data, secure multi-party computation (SMPC) enables computations on encrypted data, and federated learning supports decentralised data training

without direct sharing. Combining these approaches can enhance access control while preserving privacy. C

3. Regulatory Challenges and Cross-Border Interoperability: Existing regulatory frameworks are primarily designed with human drivers in mind and still do not fully adequately address the unique challenges posed by AVs, particularly in terms of data privacy, cybersecurity, liability and data ownership. With AV technology rapidly evolving, new regulations must be enacted to address these issues effectively. There is also an urgent need for international collaboration to establish standards that maintain and ensure data security and privacy in different regions.
4. Safety and Security Risks: AV systems face growing cybersecurity threats, requiring a multi-layered defence. Approaches such as Intrusion Detection Systems (IDS) and anomaly detection use AI-driven techniques to identify real-time cyber threats. In addition, Quantum-safe cryptography is crucial for protecting AV systems against future quantum computing attacks. Furthermore, there is an urgent need for threat prevention frameworks that incorporate risk assessment models and game theory to help proactively further mitigate potential cyber risks. By integrating these technologies, AV systems can strengthen security measures while ensuring safe and efficient driving operations.

Future work should focus on collaborative efforts among industry stakeholders, regulators, and researchers to build scalable, compliant, and secure data authorisation and validation frameworks for AVs.

Acknowledgements

The authors thank the support of the Saudi Arabian Cultural Bureau, Shaqra University, the University of York, and the EPSRC.

References

[1] A. Papadoulis, M. Quddus, M. Imprialou, Evaluating the safety impact of connected and autonomous vehicles on motorways. Accident Analysis & Prevention **124**, 12–22 (2019)

[2] S. Pan, L.M. Fulton, A. Roy, J. Jung, Y. Choi, H.O. Gao, Shared Use of Electric Autonomous Vehicles: Air Quality and Health Impacts of Future Mobility in the United States. Renewable and Sustainable Energy Reviews **149**, 111380 (2021)

[3] C. Winston, Q. Karpilow, *Autonomous vehicles: The road to economic growth?* (Brookings Institution Press, 2020)

[4] L.M. Clements, K.M. Kockelman, Economic effects of automated vehicles. Transportation research record **2606**(1), 106–114 (2017)

[5] K. Othman, Exploring the implications of autonomous vehicles: A comprehensive review. Innovative Infrastructure Solutions **7**(2), 165 (2022)

[6] S. Liu, The business case for infrastructure-vehicle cooperative autonomous driving. IEEE Engineering Management Review **50**(2), 189–194 (2022)

[7] Department for Transport. Code of Practice: Automated Vehicle Trialling (2019). URL https://www.gov.uk/government/publications/trialling-automated-vehicle-technologies-in-public/code-of-practice-automated-vehicle-trialling. Accessed: 2025-04-02

[8] D. for Transport. Rules on safe use of automated vehicles on gb roads (2022). URL https://www.gov.uk/government/consultations/safe-use-rules-for-automated-vehicles-av/rules-on-safe-use-of-automated-vehicles-on-gb-roads#consultation-on-the-rules-on-use-for-automated-lane-keeping-systems. Accessed: 2 April 2024

[9] M. Hataba, A. Sherif, M. Mahmoud, M. Abdallah, W. Alasmary, Security and privacy issues in autonomous vehicles: A layer-based survey. IEEE Open Journal of the Communications Society **3**, 811–829 (2022). https://doi.org/10.1109/OJCOMS.2022.3169500

[10] T. Alam, Data privacy and security in autonomous connected vehicles in smart city environment. Big Data and Cognitive Computing **8**(9), 95 (2024)

[11] I. Pali, R. Amin, M. Abdussami, Autonomous vehicle security: Current survey and future research challenges. Security and Privacy **7**(3), e367 (2024)

[12] Y. Xu, J. Wei, T. Mi, Z. Chen, Data security in autonomous driving: Multi-faceted challenges of technology, law, and social ethics. World Electric Vehicle Journal **16**(1), 6 (2024)

[13] F. Concas, J.K. Nurminen, T. Mikkonen, S. Tarkoma, Validation frameworks for self-driving vehicles: A survey. Smart Cities: A Data Analytics Perspective pp. 197–212 (2021)

[14] D. Omeiza, H. Webb, M. Jirotka, L. Kunze, Explanations in autonomous driving: A survey. IEEE Transactions on Intelligent Transportation Systems **23**(8), 10142–10162 (2021)

[15] N. Rajabli, F. Flammini, R. Nardone, V. Vittorini, Software Verification and Validation of Safe Autonomous Cars: A Systematic Literature Review. IEEE Access **9**, 4797–4819 (2021). https://doi.org/10.1109/ACCESS.2020.3048047

[16] R. Rajamani, *Vehicle dynamics and control* (Springer Science & Business Media, 2011)

[17] D.A. Pomerleau, *Neural network perception for mobile robot guidance*, vol. 239 (Springer Science & Business Media, 2012)

[18] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, et al., Stanley: The robot that won the darpa grand challenge. The 2005 DARPA grand challenge: the great robot race pp. 1–43 (2007)

[19] C. Urmson, J. Anhalt, D. Bagnell, C. Baker, R. Bittner, M. Clark, J. Dolan, D. Duggins, T. Galatali, C. Geyer, et al., Autonomous driving in urban environments: Boss and the urban challenge. Journal of field Robotics **25**(8), 425–466 (2008)

[20] I. Tesla, Autopilot announcement & system overview. Tech. rep., Tesla, Inc. (2015)

[21] N.H.T.S. Administration, Automated vehicles comprehensive plan. Tech. rep., National Highway Traffic Safety Administration (2020)

[22] Waymo, Waymo's fully autonomous ride-hailing service. Tech. rep., Waymo (2022). URL https://waymo.com/

[23] Baidu Apollo, Baidu's robotaxi service launch. Tech. rep., Baidu Apollo (2022). URL https://www.apollo.auto/apollo-self-driving. Accessed: April 2024

[24] E. Commission, Proposal for regulation on automated vehicles. Tech. rep., European Commission (2023). URL https://ec.europa.eu/commission/. Accessed: April 2024

[25] N.H.T.S. Administration, Automated vehicles guidance update. Tech. rep., National Highway Traffic Safety Administration (2023)

[26] UK Government, Automated vehicles bill. Tech. rep., UK Government (2024). URL https://bills.parliament.uk/bills/3506

[27] N.H.T.S. Administration, Av safety and transparency evaluation program (av step). Tech. rep., National Highway Traffic Safety Administration (2025)

[28] D.P. Watson, D.H. Scheidt, Autonomous systems. Johns Hopkins APL technical digest **26**(4), 368–376 (2005)

[29] E. Thorn, S.C. Kimmel, M. Chaka, B.A. Hamilton, et al., A framework for automated driving system testable cases and scenarios. Tech. rep., United States Department of Transportation, National Highway Traffic Safety Administration (2018). Report No. DOT HS 812 623

[30] SAE International, Sae levels of driving automation™ refined for clarity and international audience. Technical report, SAE International (2021). URL https://www.sae.org/blog/sae-j3016-update. Updated definitions of driving automation levels (SAE J3016)

[31] J. Betz, H. Zheng, A. Liniger, U. Rosolia, P. Karle, M. Behl, V. Krovi, R. Mangharam, Autonomous vehicles on the edge: A survey on autonomous vehicle racing. IEEE Open Journal of Intelligent Transportation Systems **3**, 458–488 (2022)

[32] O. Sharma, N.C. Sahoo, N. Puhan, Recent advances in motion and behavior planning techniques for software architecture of autonomous vehicles: A state-of-the-art survey. Engineering applications of artificial intelligence **101**, 104211 (2021)

[33] D.J. Yeong, G. Velasco-Hernandez, J. Barry, J. Walsh, et al., Sensor and sensor fusion technology in autonomous vehicles: A review. Sensors **21**(6), 2140 (2021)

[34] J. Vargas, S. Alsweiss, O. Toker, R. Razdan, J. Santos, An overview of autonomous vehicles sensors and their vulnerability to weather conditions. Sensors **21**(16), 5397 (2021)

[35] S.D. Pendleton, H. Andersen, X. Du, X. Shen, M. Meghjani, Y.H. Eng, D. Rus, M.H. Ang, Perception, planning, control, and coordination for autonomous vehicles. Machines **5**(1), 6 (2017)

[36] T.G. Reid, S.E. Houts, R. Cammarata, G. Mills, S. Agarwal, A. Vora, G. Pandey, Localization requirements for autonomous vehicles. arXiv preprint arXiv:1906.01061 (2019)

[37] C. Badue, R. Guidolini, R.V. Carneiro, P. Azevedo, V.B. Cardoso, A. Forechi, L. Jesus, R. Berriel, T.M. Paixao, F. Mutz, et al., Self-driving Cars: A survey. Expert systems with applications **165**, 113816 (2021)

[38] J. Zhao, W. Zhao, B. Deng, Z. Wang, F. Zhang, W. Zheng, W. Cao, J. Nan, Y. Lian, A.F. Burke, Autonomous Driving System: A Comprehensive Survey. Expert Systems with Applications p. 122836 (2023)

[39] J. Fayyad, M.A. Jaradat, D. Gruyer, H. Najjaran, Deep learning sensor fusion for autonomous vehicle perception and localization: A review. Sensors **20**(15), 4220 (2020)

[40] M.N. Ahangar, Q.Z. Ahmed, F.A. Khan, M. Hafeez, A Survey of Autonomous Vehicles: Enabling Communication Technologies and Challenges. Sensors **21**(3), 706 (2021)

[41] T. Nguyen, M. Yoo, *Fusing LIDAR sensor and RGB camera for object detection in autonomous vehicle with fuzzy logic approach*, in *2021 International Conference on Information Networking (ICOIN)* (IEEE, 2021), pp. 788–791

[42] S. Campbell, N. O'Mahony, L. Krpalcova, D. Riordan, J. Walsh, A. Murphy, C. Ryan, *Sensor technology in autonomous vehicles: A review*, in *2018 29th Irish*

Signals and Systems Conference (ISSC) (IEEE, 2018), pp. 1–4

[43] J. Wishart, S. Como, U. Forgione, J. Weast, et al., Literature review of verification and validation activities of automated driving systems. SAE International Journal of Connected and Automated Vehicles **3**(4), 267–323 (2020)

[44] M. Martínez-Díaz, F. Soriguera, Autonomous vehicles: theoretical and practical challenges. Transportation Research Procedia **33**, 275–282 (2018)

[45] Z. Szalay, D. Ficzere, V. Tihanyi, F. Magyar, G. Soós, P. Varga, 5G-enabled Autonomous Driving Demonstration with a V2X Scenario-In-The-Loop Approach. Sensors **20**(24), 7344 (2020)

[46] S. Guleng, C. Wu, Z. Liu, X. Chen, Edge-based V2X Communications with Big Data Intelligence. IEEE Access **8**, 8603–8613 (2020)

[47] V. Kulshrestha, K.R. Jagdale, in *Towards Wireless Heterogeneity in 6G Networks* (CRC Press, 2024), pp. 18–32

[48] M. Noor-A-Rahim, Z. Liu, H. Lee, M.O. Khyam, J. He, D. Pesch, K. Moessner, W. Saad, H.V. Poor, 6G for Vehicle-to-Everything (V2X) Communications: Enabling Technologies, Challenges, and Opportunities. Proceedings of the IEEE **110**(6), 712–734 (2022)

[49] N.R. Fulbright. The privacy implications of autonomous vehicles (2017)

[50] M. Cebe, E. Erdin, K. Akkaya, H. Aksu, S. Uluagac, Block4Forensic: An Integrated Lightweight Blockchain Framework for Forensics Applications of Connected Vehicles. IEEE Communications Magazine **56**(10), 50–57 (2018)

[51] J. Guanetti, Y. Kim, F. Borrelli, Control of Connected and Automated Vehicles: State of The Art and Future Challenges. Annual reviews in control **45**, 18–40 (2018)

[52] Y. Zhang, C.G. Cassandras, An Impact Study of Integrating Connected Automated Vehicles with Conventional Traffic. Annual Reviews in Control **48**, 347–356 (2019)

[53] P. Yi. Access control at major/minor road intersection through cav in mixed traffic. https://rosap.ntl.bts.gov/view/dot/73985 (2024). Government report, Accessed: 2024-05-13

[54] J. Ying, Y. Feng, Infrastructure-assisted cooperative driving and intersection management in mixed traffic conditions. Transportation Research Part C: Emerging Technologies **158**, 104443 (2024)

[55] D.C. Marinescu, *Cloud Computing: Theory and Practice* (Morgan Kaufmann, 2022)

[56] S. Sharma, Nidhi, *Vehicular ad-hoc network: An overview*, in *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)* (IEEE, 2019), pp. 131–134. https://doi.org/10.1109/ICCCIS48478.2019.8974524

[57] K. Stefanovic, M. Malnar, N. Jevtic, *An Overview of Communication Technologies for VANET*, in *2024 23rd International Symposium INFOTEH-JAHORINA (INFOTEH)* (IEEE, 2024), pp. 1–6

[58] J. He, K. Yang, H.H. Chen, 6G Cellular Networks and Connected Autonomous Vehicles. IEEE Network **35**(4), 255–261 (2020)

[59] M.A. Khan, H.E. Sayed, S. Malik, T. Zia, J. Khan, N. Alkaabi, H. Ignatious, Level-5 Autonomous Driving—Are We There Yet? A Review of Research Literature. ACM Computing Surveys (CSUR) **55**(2), 1–38 (2022)

[60] The 5G Infrastructure Public Private Partnership (5G PPP). URL http://www.5g-ppp.eu

[61] S. Hakak, T.R. Gadekallu, S.P. Ramu, P.K.R. Maddikunta, C. de Alwis, M. Liyanage, et al., Autonomous vehicles in 5g and beyond: A survey. arXiv preprint arXiv:2207.10510 (2022)

[62] D. Srivastava, P. Sahu, G. Geetha, *Autonomous ground vehicle for physically impaired: A solution rather than luxury*, in *2022 International Conference on Computer Communication and Informatics (ICCCI)* (IEEE, 2022), pp. 1–11

[63] C.D. Harper, C.T. Hendrickson, C. Samaras, Cost and Benefit Estimates of Partially-Automated Vehicle Collision Avoidance Technologies. Accident Analysis & Prevention **95**, 104–115 (2016)

[64] Net Zero by 2050: A Roadmap for the Global Energy Sector (2021). URL https://trid.trb.org/view/1856381

[65] X. Sun, F.R. Yu, P. Zhang, A survey on Cyber-Security of Connected and Autonomous Vehicles (CAVs). IEEE Transactions on Intelligent Transportation Systems (2021)

[66] V.C. Hu, D. Ferraiolo, R. Kuhn, A.R. Friedman, A.J. Lang, M.M. Cogdell, A. Schnitzer, K. Sandlin, R. Miller, K. Scarfone, et al., Guide to Attribute Based Access Control (abac) Definition and Considerations (draft). NIST special publication **800**(162), 1–54 (2013)

[67] R.S. Sandhu, in *Advances In Computers*, vol. 46 (Elsevier, 1998), pp. 237–286

[68] J. Park, R. Sandhu, The UCONABC Usage Control Model. ACM Transactions on Information and System Security (TISSEC) **7**(1), 128–174 (2004)

[69] K. Christidis, M. Devetsikiotis, Blockchains and Smart Contracts for the Internet of Things. IEEE access **4**, 2292–2303 (2016)

[70] C. Wolter, C. Meinel, An Approach to Capture Authorisation Requirements in Business Processes. Requirements Engineering **15**, 359–373 (2010)

[71] A. Khan, A. Ahmad, M. Ahmed, J. Sessa, M. Anisetti, Authorization Schemes for Internet of Things: Requirements, Weaknesses, Future Challenges and Trends. Complex & Intelligent Systems **8**(5), 3919–3941 (2022)

[72] A. Arooj, M.S. Farooq, A. Akram, R. Iqbal, A. Sharma, G. Dhiman, Big Data Processing and Analysis in Internet of Vehicles: Architecture, Taxonomy, and Open Research Challenges. Archives of Computational Methods in Engineering **29**(2), 793–829 (2022)

[73] L. Liu, S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, W. Shi, Computing Systems for Autonomous Driving: State of the Art and Challenges. IEEE Internet of Things Journal **8**(8), 6469–6486 (2020)

[74] J. Ryu, Y. Lee, Y. Yoon, *Blockchain Model for Reliable Consensus Algorithm on the Autonomous Driving Data Management*, in *2024 IEEE International Conference on Consumer Electronics (ICCE)* (2024), pp. 1–4. https://doi.org/10.1109/ICCE59016.2024.10444480

[75] R. Changalvala, Sensor Data Integrity Verification for Real-time and Resource Constrained Systems. Ph.D. thesis, University of Michigan Library (2021)

[76] M.A. Javed, M.Z. Khan, U. Zafar, M.F. Siddiqui, R. Badar, B.M. Lee, F. Ahmad, ODPV: An Efficient Protocol to Mitigate Data Integrity Attacks in Intelligent Transport Systems. IEEE Access **8**, 114733–114740 (2020)

[77] X. Shen, Y. Lu, Y. Zhang, X. Liu, L. Zhang, An Innovative Data Integrity Verification Scheme in the Internet of Things Assisted information Exchange in Transportation Systems. Cluster Computing **25**(3), 1791–1803 (2022)

[78] A. Rakhmanov, Y. Wiseman, Compression of GNSS Data with the Aim of Speeding up Communication to Autonomous Vehicles. Remote Sensing **15**(8), 2165 (2023). https://doi.org/10.3390/rs15082165. URL https://doi.org/10.3390/rs15082165

[79] H. Min, Y. Fang, X. Wu, X. Lei, S. Chen, R. Teixeira, B. Zhu, X. Zhao, Z. Xu, A Fault Diagnosis Framework for Autonomous Vehicles with Sensor Self-Diagnosis. Expert Systems with Applications **224**, 120002 (2023)

[80] S. Chuprov, I. Viksnin, I. Kim, L. Reznikand, I. Khokhlov, *Reputation and Trust Models with Data Quality Metrics for Improving Autonomous Vehicles Traffic Security and Safety*, in *2020 IEEE Systems Security Symposium (SSS)* (2020),

pp. 1–8. https://doi.org/10.1109/SSS47320.2020.9174269

[81] Tesla, Inc. Electric cars (2022). URL https://www.tesla.com/en_gb. Accessed: 2024-05-13

[82] J.T. Correia, K.A. Iliadis, E.S. McCarron, M.A. Smolej, B. Hastings, C.C. Engineers, *Utilizing data from automotive event data recorders*, in *Proceedings of the Canadian Multidisciplinary Road Safety Conference XII, London Ontario* (2001), p. 18

[83] International Organization of Motor Vehicle Manufacturers. URL https://www.oica.net

[84] K. Böhm, T. Kubjatko, D. Paula, H.G. Schweiger, New Developments on EDR (Event Data Recorder) for Automated Vehicles. Open Engineering **10**(1), 140–146 (2020)

[85] Responsible Innovation in Self-Driving Vehicles. https://www.gov.uk/government/publications/responsible-innovation-in-self-driving-vehicles/responsible-innovation-in-self-driving-vehicles (2022). Accessed: 2022

[86] RIMKUS. Event Data Recorder Supported Vehicles. URL https://rimkus.com/media/pdfs/Event-Data-Recorder-Vehicle-List-Rimkus-1.11.21.pdf. Accessed: May 2, 2022

[87] J. Gwehenberger, O. Braxmeier, C. Lauterwasser, M.A. Kreutner, M. Borrack, C. Reinkemeyer, L. Wech, M. Weyde, P. Salzberger, Needs and Requirements of EDR for Automated Vehicles: Analysis Based on Insurance Claims Reported to Allianz Germany. Tech. rep., Allianz Center for Technology (AZT) (2022). URL https://unece.org/sites/default/files/2022-03/03_EDR_Paper_Allianz_Germany.pdf. Presented at UNECE EDR Workshop, March 2022

[88] EU Approval and Market Surveillance Measures for Motor Vehicles and Their Trailers. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:4350062

[89] European Union. European Union Official Website (2024). URL https://european-union.europa.eu/index_en. Accessed: 2024-10-06

[90] World-wide level. URL https://www.connectedautomateddriving.eu/regulation-and-policies/world-wide-harmonization/

[91] N.H.T.S. Administration. Press release: Nhtsa-ihhs commitment on aeb. http://www.nhtsa.gov/About+NHTSA (2015). Accessed: 2022

[92] D. Paula, K. Böhm, T. Kubjatko, H.G. Schweiger, *Challenges in Forensic Reconstruction of Traffic Accidents Involving Advanced Driver Assistance Systems*

(ADAS), in *29th Annual Congress of the European Association for Accident Research (EVU)* (2020)

[93] P. Koopman, M. Wagner, Autonomous Vehicle Safety: An Interdisciplinary Challenge. IEEE Intelligent Transportation Systems Magazine **9**(1), 90–96 (2017). https://doi.org/10.1109/MITS.2016.2583491

[94] Review of the Existing National / Regional Activities and a Proposed Way Forward for EDR (2022). URL https://unece.org/transport/documents/2022/03/informal-documents/edrdssa-iwg-review-existing-national-regional

[95] I. Kim, G. Lee, S. Lee, W. Choi, *Cybersecurity and Capacity Requirement for Data Storage of Autonomous Driving System*, in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)* (IEEE, 2022), pp. 1–7

[96] I. Kim, G. Lee, S. Lee, W. Choi, *Data Storage System Requirement for Autonomous Vehicle*, in *2022 22nd International Conference on Control, Automation and Systems (ICCAS)* (IEEE, 2022), pp. 45–49

[97] C. Ten Holter, L. Kunze, J.A. Pattinson, P. Salvini, J. Attias, M. Jirotka, *What's missing from this picture? Ethical, legal, and practical challenges for autonomous-vehicle data-recorders*, in *Proceedings of the Second International Symposium on Trustworthy Autonomous Systems* (2024), pp. 1–13

[98] R. Patriarca, G. Di Gravio, R. Woltjer, F. Costantino, G. Praetorius, P. Ferreira, E. Hollnagel, Framing the FRAM: A literature Review on the Functional Resonance Analysis Method. Safety Science **129**, 104827 (2020)

[99] M.A. Hoque, R. Hasan, *AVGuard: A Forensic Investigation Framework for Autonomous Vehicles*, in *ICC 2021-IEEE International Conference on Communications* (IEEE, 2021), pp. 1–6

[100] C. Oham, S.S. Kanhere, R. Jurdak, S. Jha, A Blockchain Based Liability Attribution Framework for Autonomous Vehicles. arXiv preprint arXiv:1802.05050 (2018)

[101] S. Lee, W. Choi, H.J. Jo, D.H. Lee, T-box: A Forensics-Enabled Trusted Automotive Data Recording Method. IEEE Access **7**, 49738–49755 (2019)

[102] K.K.G. Buquerin, C. Corbett, H.J. Hof, A Generalized Approach to Automotive Forensics. Forensic Science International: Digital Investigation **36**, 301111 (2021)

[103] W. Liu, W. Shen, L. Harn, M. Luo, A Fast VANET-Assisted Scheme for Event Data Recorders. Security and Communication Networks **2022** (2022)

[104] J. Beck, R. Arvin, S. Lee, A. Khattak, S. Chakraborty, Automated Vehicle Data Pipeline for Accident Reconstruction: New Insights from LiDAR, Camera, and

Radar Data. Accident Analysis & Prevention **180**, 106923 (2023)

[105] A. Singh, S. Sural, T. Sengupta, S. Sural, *Trusted Sharing of Autonomous Vehicle Crash Data using Enterprise Blockchain and IPFS*, in *Proceedings of the 5th ACM International Symposium on Blockchain and Secure Critical Infrastructure* (2023), pp. 11–24

[106] A. Elhadeedy, J. Daily, *Autonomous Vehicle Development as a System of Systems and Constituent Systems Networking*, in *2024 IEEE International Systems Conference (SysCon)* (IEEE, 2024), pp. 1–7

[107] S. Sonko, E.A. Etukudoh, K.I. Ibekwe, V.I. Ilojianya, C.D. Daudu, A Comprehensive Review of Embedded Systems in Autonomous Vehicles: Trends, Challenges, and Future Directions. World Journal of Advanced Research and Reviews **21**(1), 2009–2020 (2024)

[108] S. Olariu, A Survey of Vehicular Cloud Research: Trends, Applications and Challenges. IEEE Transactions on Intelligent Transportation Systems **21**(6), 2648–2663 (2019)

[109] J. Kang, D. Lin, E. Bertino, O. Tonguz, *From Autonomous Aehicles to Vehicular Clouds: Challenges of Management, security and Dependability*, in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)* (IEEE, 2019), pp. 1730–1741

[110] A. Ghansiyal, M. Mittal, A.K. Kar, Information Management Challenges in Autonomous Vehicles: A Systematic Literature Review. Journal of Cases on Information Technology (JCIT) **23**(3), 58–77 (2021)

[111] T.X. Tran, D. Pompili, Joint Task Offloading and Resource Allocation for Multi-Server Mobile-Edge Computing Networks. IEEE Transactions on Vehicular Technology **68**(1), 856–868 (2018)

[112] Y. Liu, H. Yu, S. Xie, Y. Zhang, Deep Reinforcement Learning for Offloading and Resource Allocation in Vehicle Edge Computing and Networks. IEEE Transactions on Vehicular Technology **68**(11), 11158–11168 (2019)

[113] J. Zhao, Q. Li, Y. Gong, K. Zhang, Computation Offloading and Resource Allocation For Cloud Assisted Mobile Edge Computing in Vehicular Networks. IEEE Transactions on Vehicular Technology **68**(8), 7944–7956 (2019). https://doi.org/10.1109/TVT.2019.2917890

[114] J. Ren, G. Yu, Y. He, G.Y. Li, Collaborative Cloud and Edge Computing for Latency Minimization. IEEE Transactions on Vehicular Technology **68**(5), 5031–5044 (2019)

[115] W. Sun, H. Zhang, R. Wang, Y. Zhang, Reducing Offloading Latency for Digital Twin Edge Networks in 6G. IEEE Transactions on Vehicular Technology **69**(10), 12240–12251 (2020)

[116] R. Bi, J. Xiong, Y. Tian, Q. Li, X. Liu, Edge-Cooperative Privacy-Preserving Object Detection Over Random Point Cloud Shares for Connected Autonomous Vehicles. IEEE Transactions on Intelligent Transportation Systems **23**(12), 24979–24990 (2022)

[117] M. Asim, A.A. Abd El-Latif, Intelligent Computational Methods for Multi-Unmanned Aerial Vehicle-Enabled Autonomous Mobile Edge Computing Systems. ISA transactions **132**, 5–15 (2023)

[118] K. Gai, M. Qiu, H. Zhao, Privacy-Preserving Data Encryption Strategy for Big Data in Mobile Cloud Computing. IEEE Transactions on Big Data **7**(4), 678–688 (2021). https://doi.org/10.1109/TBDATA.2017.2705807

[119] J. Liang, J. Zhang, V.C. Leung, X. Wu, Distributed Information Exchange with Low Latency for Decision Making in Vehicular Fog Computing. IEEE Internet of Things Journal (2021)

[120] SAE International. Autonomous Vehicles and Their Cloud-Computing Networks (2021). URL https://www.sae.org/news/2021/04/autonomous-vehicles-and-their-cloud-computing-networks

[121] Autonomous Vehicle Data Storage (2020). URL https://blocksandfiles.com/2020/02/03/autonomous-vehicle-data-storage-is-a-game-of-guesses/

[122] Cars in the Clouds (2021). URL https://www.rackwareinc.com/driving-with-the-cloud

[123] Z. Wang, Y. Wu, Q. Niu, Multi-Sensor Fusion in Automated Driving: A survey. Ieee Access **8**, 2847–2868 (2019)

[124] D. Geradin, Access to In-Vehicle Data by Third-Party Service Providers: Is there a Market Failure and, if so, How Should it be Addressed? How Should it be Addressed (2020)

[125] M. McCarthy, M. Seidl, S. Mohan, J. Hopkin, A. Stevens, F. Ognissanto, Access to In-Vehicle Data and Resources. Study comissioned by European Commission CPR2419. Brussels p. 10 (2017)

[126] S. Singh, Critical reasons for crashes investigated in the national motor vehicle crash causation survey. Traffic Safety Facts - Crash Stats DOT HS 812 115, National Center for Statistics and Analysis, National Highway Traffic Safety Administration (NHTSA), Washington, DC (2015). URL http://www-nrd.nhtsa.dot.gov/Pubs/812115.pdf. Open Access report, includes tables

[127] Q. Yuan, Y. Peng, X. Xu, X. Wang, Key Points of Investigation and Analysis on Traffic Accidents Involving Intelligent Vehicles. Transportation Safety and Environment **3**(4), tdab020 (2021)

[128] D.C. Nguyen, M. Ding, P.N. Pathirana, A. Seneviratne, J. Li, H.V. Poor, Federated Learning for Internet of Things: A Comprehensive Survey. IEEE Communications Surveys & Tutorials **23**(3), 1622–1658 (2021)

[129] D. Byrd, A. Polychroniadou, *Differentially Private Secure Multi-Party Computation for Federated Learning in Financial Applications*, in *Proceedings of the First ACM International Conference on AI in Finance* (2020), pp. 1–9

[130] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H.B. McMahan, S. Patel, D. Ramage, A. Segal, K. Seth, *Practical Secure Aggregation for Privacy-Preserving Machine Learning*, in *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (2017), pp. 1175–1191

[131] D. Froelicher, J.R. Troncoso-Pastoriza, J.L. Raisaro, M.A. Cuendet, J.S. Sousa, H. Cho, B. Berger, J. Fellay, J.P. Hubaux, Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multiparty Homomorphic Encryption. Nature communications **12**(1), 1–10 (2021)

[132] E. Sotthiwat, L. Zhen, Z. Li, C. Zhang, *Partially Encrypted Multi-Party Computation for Federated Learning*, in *2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid)* (IEEE, 2021), pp. 828–835

[133] S.R. Pokhrel, J. Choi, Federated learning with blockchain for autonomous vehicles: Analysis and design challenges. IEEE Transactions on Communications **68**(8), 4734–4746 (2020)

[134] N.H. Nguyen, T.A. Nguyen, T. Nguyen, V.T. Hoang, D.D. Le, K.S. Wong, *Towards Efficient Communication and Secure Federated Recommendation System via Low-rank Training*, in *Proceedings of the ACM on Web Conference 2024* (2024), pp. 3940–3951

[135] T. Zeng, O. Semiariy, M. Chen, W. Saad, M. Bennis, Federated Learning on the Road Autonomous Controller Design for Connected and Autonomous Vehicles. IEEE Transactions on Wireless Communications (2022)

[136] S. Liu, Y. Shang, *Secure Resource Sharing on Hyperledger Fabric based on CP-ABE*, in *2021 The 3rd International Conference on Blockchain Technology* (2021), pp. 203–209

[137] X. Zhao, S. Wang, Y. Zhang, Y. Wang, Attribute Based Access Control Scheme for Data Sharing on Hyperledger Fabric. Journal of Information Security and Applications **67**, 103182 (2022)

[138] G. Bianchi, T. Dargahi, A. Caponi, M. Conti, Intelligent Conditional Collaborative Private Data Sharing. Future Generation Computer Systems **96**, 1–10 (2019)

[139] R. Alhabib, P. Yadav, *Hyperledger Fabric Platform for Secure and Efficient Data Sharing in Autonomous Vehicles*, in *University of York, UK* (IEEE Communications Society, 2024)

[140] J.A. Khan, W. Wang, K. Ozbay, BELIEVE: Privacy-Aware Secure Multi-Party Computation for Real-Time Connected and Autonomous Vehicles and Micro-Mobility Data Validation Using Blockchain—A Study on New York City Data. Transportation research record **2678**(3), 410–421 (2024)

[141] V. Pillai, Techniques for Processing and Analyzing Large Data Sets Using Big Data Analytics. Available at SSRN 4991707 (2024)

[142] X. Zuo, *Research on Data Quality Improvement Program Based on Big Data Application*, in *2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, vol. 3 (IEEE, 2023), pp. 1742–1745

[143] P.H. Ahmad, M. Rai, *Analysis of Optimization Strategies for Big Data Storage Management: A Study*, in *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)* (IEEE, 2023), pp. 1747–1753

[144] X. Zhang, J. Tao, K. Tan, M. Törngren, J.M.G. Sánchez, M.R. Ramli, X. Tao, M. Gyllenhammar, F. Wotawa, N. Mohan, et al., Finding critical scenarios for automated driving systems: A systematic literature review. arXiv preprint arXiv:2110.08664 (2021)

[145] C. Neurohr, L. Westhofen, M. Butz, M.H. Bollmann, U. Eberle, R. Galbas, Criticality Analysis for the Verification and Validation of Automated Vehicles. IEEE Access **9**, 18016–18041 (2021)

[146] X. Zhao, Y. Fang, H. Min, X. Wu, W. Wang, R. Teixeira, Potential Sources of Sensor Data Anomalies for Autonomous Vehicles: An Overview from Road Vehicle Safety Perspective. Expert Systems with Applications **236**, 121358 (2024)

[147] A. Kamel, T. Sayed, M. Kamel, Real-time Combined Safety-Mobility Assessment Using Self-Driving Vehicles Collected Data. Accident Analysis & Prevention **199**, 107513 (2024)

[148] L. Rylander, J. Englund, Conceptual Fault-Handling System Design for Driverless Trucks A case Study Based on Industry Practices in Sweden. Transportation research interdisciplinary perspectives **25**, 101123 (2024)

[149] P. Koopman, M. Wagner, Challenges in Autonomous Vehicle Testing and Validation. SAE International Journal of Transportation Safety **4**(1), 15–24 (2016)

[150] G. Griessnig, A. Schnellbach, *Development of the 2nd Edition of the ISO 26262*, in *European Conference on Software Process Improvement* (Springer, 2017), pp. 535–546

[151] F. Henao. Why Data Handling May Put a Bump on the Road to Autonomous Driving. URL https://europe.autonews.com/guest-columnist/why-data-handling-may-put-bump-road-autonomous-driving

[152] R.G. Goriparthi, AI-driven predictive analytics for autonomous systems: A machine learning approach. Revista de Inteligencia Artificial en Medicina **15**(1), 843–879 (2024)

[153] A. Sultan, S. Tahir, H. Tahir, T. Anwer, F. Khan, M. Rajarajan, O. Rana, A Novel Image-Based Homomorphic Approach for Preserving the Privacy of Autonomous Vehicles Connected to the Cloud. IEEE Transactions on Intelligent Transportation Systems **24**(2), 1936–1948 (2023). https://doi.org/10.1109/TITS.2022.3219591

[154] Y. Zhang, J. Zou, R. Guo, Efficient privacy-preserving authentication for V2G networks. Peer-to-Peer Networking and Applications **14**(3), 1366–1378 (2021)

[155] R. Parekh, N. Patel, R. Gupta, N.K. Jadav, S. Tanwar, A. Alharbi, A. Tolba, B.C. Neagu, M.S. Raboaca, Gefl: Gradient Encryption-Aided Privacy Preserved Federated Learning for Autonomous Vehicles. IEEE Access **11**, 1825–1839 (2023)

[156] M. Gheisari, W.Z. Khan, H.E. Najafabadi, G. McArdle, H. Rabiei-Dastjerdi, Y. Liu, C. Fernández-Campusano, H.B. Abdalla, CAPPAD: a Privacy-Preservation Solution for Autonomous Vehicles using SDN, Differential Privacy and Data Aggregation. Applied Intelligence **54**(4), 3417–3428 (2024)

[157] T. Bai, Q. Yang, S. Fu, *User-Defined Privacy Preserving Data Sharing for Connected Autonomous Vehicles Utilizing Edge Computing*, in *Proceedings of the Eighth ACM/IEEE Symposium on Edge Computing* (Association for Computing Machinery, New York, NY, USA, 2024), SEC '23, p. 145–157. https://doi.org/10.1145/3583740.3628436. URL https://doi.org/10.1145/3583740.3628436

[158] Y. Zhao, D. Gong, S. Wen, L. Ding, G. Guo, A Privacy-Preserving-Based Distributed Collaborative Scheme for Connected Autonomous Vehicles at Multi-Lane Signal-Free Intersections. IEEE Transactions on Intelligent Transportation Systems (2024)

[159] S. Gupta, C. Maple. A Survey of Security Mechanisms for Edge Computing Based Connected Autonomous Vehicles (2022). URL https://techrxiv.org/. Preprint published in TechRxiv

[160] J. Li, Z. Xue, C. Li, M. Liu, RTED-SD: A Real-Time Edge Detection Scheme for Sybil DDoS in the Internet of Vehicles. IEEE Access **9**, 11296–11305 (2021)

[161] A. Anwar, T. Halabi, M. Zulkernine, A Dynamic Threat Prevention Framework for Autonomous Vehicle Networks Based on Ruin-Theoretic Security Risk Assessment. Journal on Autonomous Transportation Systems **1**(4), 1–28 (2024)

[162] S. Aurangzeb, M. Aleem, M.T. Khan, H. Anwar, M.S. Siddique, Cybersecurity for Autonomous Vehicles Against Malware Attacks in Smart-Cities. Cluster Computing **27**(3), 3363–3378 (2024)

[163] A.M. Cretu, F. Houssiau, A. Cully, Y.A. de Montjoye, *QuerySnout: Automating the Discovery of Attribute Inference Attacks Against Query-based Systems*, in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security* (2022), pp. 623–637

[164] C. Anthony, W. Elgenaidi, M. Rao, Intrusion Detection System for Autonomous Vehicles Using Non-Tree based Machine Learning Algorithms. Electronics **13**(5), 809 (2024)

[165] S. Anbalagan, G. Raja, S. Gurumoorthy, R.D. Suresh, K. Dev, IIDS: Intelligent Intrusion Detection System for Sustainable Development in Autonomous Vehicles. IEEE Transactions on Intelligent Transportation Systems **24**(12), 15866–15875 (2023)

[166] F. Aloraini, A. Javed, O. Rana, Adversarial Attacks on Intrusion Detection Systems in In-Vehicle Networks of Connected and Autonomous Vehicles. Sensors **24**(12), 3848 (2024)

[167] H. Setia, A. Chhabra, S.K. Singh, S. Kumar, S. Sharma, V. Arya, B.B. Gupta, J. Wu, Securing the Road Ahead: Machine Learning-Driven DDoS Attack Detection in VANET Cloud Environments. Cyber Security and Applications **2**, 100037 (2024)

[168] D. Chaudhary, P. Santhi, M.S.P. Durgarao, A. Padmavathi, M. Mehedi Hassan, B. Fahad Alkhamees, Module Lattice-Based Post Quantum Secure Blockchain Empowered Authentication Framework for Autonomous Truck Platooning. IEEE Access **12**, 105219–105233 (2024)

[169] K. Sutradhar, A Quantum Cryptographic Protocol for Secure Vehicular Communication. IEEE Transactions on Intelligent Transportation Systems **25**(5), 3513–3522 (2024). https://doi.org/10.1109/TITS.2023.3322728

[170] Q. Xu, L. Zhao, Z. Su, D. Fang, R. Li, Secure Federated Learning in Quantum Autonomous Vehicular Networks. IEEE Network **37**(6), 240–247 (2023). https://doi.org/10.1109/MNET.134.2200619

[171] S.K. Dwivedi, R. Amin, S. Vollala, R. Chaudhry, Blockchain-based Secured Event-Information Sharing Protocol in Internet of Vehicles for Smart Cities. Computers & Electrical Engineering **86**, 106719 (2020)

[172] H. Yi, A Secure Blockchain System for Internet of Vehicles based on 6G-Enabled Network in Box. Computer Communications **186**, 45–50 (2022)

[173] G. Liu, H. Dong, Z. Yan, X. Zhou, S. Shimizu, B4SDC: A blockchain System for Security Data Collection in MANETs. IEEE Transactions on Big Data (2020)

[174] A. Faisal, M. Kamruzzaman, T. Yigitcanlar, G. Currie, Understanding Autonomous Vehicles. Journal of transport and land use **12**(1), 45–72 (2019)

[175] Connected and Automated mobility (2022). URL https://digital-strategy.ec.europa.eu/en/policies/connected-and-automated-mobility#:~:text=Connected%20and%20Automated%20Mobility%20(CAM,guide%20themselves%20without%20human%20intervention)

[176] D.B. Resnik, S.L. Andrews, A Precautionary Approach to Autonomous Vehicles. AI and Ethics **4**(2), 403–418 (2024)

[177] T. Pöysti, The Precautionary Approach Design Pattern. Digital Society **3**(1), 5 (2024)

[178] R. Horne, C. Law-Walsh, Z. Assaad, K. Joiner, *Ten Regulatory Principles to Scaffold the Design, Manufacture, and Use of Trustworthy Autonomous Systems, Illustrated in a Maritime Context*, in *Proceedings of the First International Symposium on Trustworthy Autonomous Systems* (Association for Computing Machinery, New York, NY, USA, 2023), TAS '23. https://doi.org/10.1145/3597512.3599701. URL https://doi.org/10.1145/3597512.3599701

[179] M. Lu, O. Turetken, O.E. Adali, J. Castells, R. Blokpoel, P. Grefen, *C-ITS (Cooperative Intelligent Transport Systems) Deployment in Europe: Challenges and Key Findings*, in *25th ITS World Congress, Copenhagen, Denmark* (2018), pp. 17–21

[180] L. Vlacic, Cross-Border, Interoperable Cooperative Intelligent Transportation Systems: Could We Make Them Operational? IEEE Intelligent Transportation Systems Magazine **14**(4), 3–4 (2022)

[181] J. Naranjo, F. Jiménez, R. Castiñeira, M. Gil, C. Premebida, P. Serra, A. Valejo, F. Nashashibi, C. Magalhães, Cross-Border Interoperability for Cooperative, Connected and Autonomous Driving. IEEE Intelligent Transportation Systems

Magazine (2021)