



Review

Sensor Fusion and Perception for Autonomous Driving: A Critical Review of Modalities, AI Models, Algorithms, and Industry Configurations

Esraa Khatab ¹ , Fares Fathy ² , Abdallah AlKholy ² and Omar Shalash ^{3,*} ¹ School of Mathematical and Computer Sciences, Heriot-Watt University, Dubai P.O. Box 501745, United Arab Emirates; e.khatab@hw.ac.uk² Arab Academy for Science, Technology and Maritime Transport, Alamein 19838, Egypt; f.ibrahim0329@student.aast.edu (F.F.); a.alkholy53@student.aast.edu (A.A.)³ Artificial Intelligence Research Center (AIRC), College of Engineering and Information Technology, Ajman University, Ajman P.O. Box 346, United Arab Emirates

* Correspondence: o.shalash@ajman.ac.ae

Abstract

Autonomous driving systems rely on a sophisticated pipeline of artificial intelligence models to perceive, predict, and plan in dynamic environments. This review presents a systematic analysis of the machine learning and deep learning models underpinning vehicle autonomy, spanning classical convolutional neural networks (CNNs) for object detection and semantic segmentation to recurrent and Transformer-based architectures for trajectory prediction and motion planning. It also provides a critical examination of the autonomous vehicle sensor stack, including cameras, LiDAR, radar, ultrasonics, and GNSS/IMU as data acquisition systems, highlighting modality-specific AI challenges such as monocular depth estimation, 3D point cloud processing, and radar Doppler interpretation. The evolution of perception and decision-making pipelines is reviewed, contrasting modular architectures with end-to-end learning paradigms that directly map raw sensor data to control commands, and discussing their trade-offs in interpretability, safety assurance, and robustness to rare edge cases. We further survey specialized hardware accelerators and heterogeneous automotive SoCs designed to meet stringent real-time and power constraints. Industrial strategies are compared, including multi-modal sensor fusion and vision-centric approaches based on large-scale imitation learning. Finally, we identify open challenges related to robustness under adverse conditions, domain shift, causal ambiguity, and the need for interpretable and certifiable AI in safety-critical autonomous driving systems.

Keywords: autonomous driving; deep learning; sensor fusion; knowledge extraction; computer vision



Academic Editor: Marek Sikora

Received: 17 April 2026

Revised: 25 June 2026

Accepted: 27 June 2026

Published: 7 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Autonomous driving systems are a major application of machine learning because they require the extraction of actionable knowledge from high-dimensional, real-time sensor data and its conversion into safe driving decisions. The field has advanced rapidly because of progress in perception, prediction, planning, and embedded computing, but reliable deployment is still limited by challenges in robustness, interpretability, and validation [1–6]. The Society of Automotive Engineers (SAE) J3016 standard provides a useful framework for describing the progression from Level 0 to Level 5 automation and for positioning current systems within this broader technological landscape, as shown in Figure 1 and Table 1.

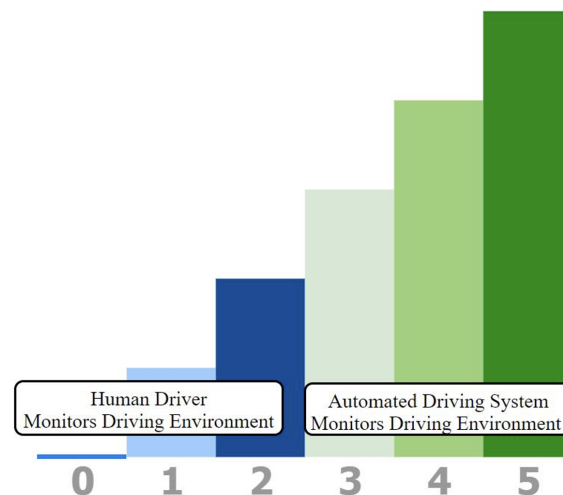


Figure 1. SAE J3016 automotive automation standard [7].

Table 1. SAE-defined Levels of Autonomous Technology. Adapted from SAE [8].

Level	Definition	Description
Level 0	No Autonomous Technology	Vehicles can help drivers with basic activities, including acceleration, braking, and steering.
Level 1	Driver Assistance	Vehicles can manage the direction or speed, but not both, while the driver must take full responsibility for driving
Level 2	Partial Automation	Vehicles may accelerate, brake, and steer; operations can be conducted simultaneously; however, the driver must maintain control of the vehicle the entire time.
Level 3	Conditional Automation	Vehicles can do autonomous driving in some conditions, but drivers must always be ready to take control
Level 4	High Automation	Vehicles are self-driving and do not require a human operator, but the driver can interfere in specific situations.
Level 5	Full Automation	Vehicle are entirely autonomous in all situations. Passengers only need to provide information about the location of the vehicle. Steering wheels and pedals are unlikely to be available

Advanced Driver Assistance Systems (ADAS) support drivers in tasks such as navigation and parking by using camera-based sensors without fully taking over the driving process. Their purpose is to reduce accidents by analyzing traffic conditions, congestion, road hazards, and other environmental factors. At the core of most ADAS platforms is a hardware accelerator responsible for interpreting the vehicle's surroundings and identifying potential risks. These systems typically rely on four main perception sensors: LiDAR, radar, cameras, and ultrasonic sensors. The data collected from these sensors are processed through the accelerator and fused to detect nearby objects, including pedestrians, vehicles, lane markings, and traffic signs, before being passed to the braking, steering, and throttle control. This overall workflow is illustrated in Figure 2 [9].

Recent advances in the field have been largely propelled by breakthroughs in deep learning models and the availability of specialized hardware accelerators [10–12]. Autonomous vehicles rely on a diverse sensor suite to capture a multi-modal representation of their environment [13,14]. Machine learning algorithms are then tasked with fusing this heterogeneous data to perform critical knowledge extraction tasks such as object detection, scene understanding, localization, and path planning. This evolution from traditional

rule-based systems to end-to-end learning paradigms marks a significant shift in how intelligent vehicles interpret and navigate the world.

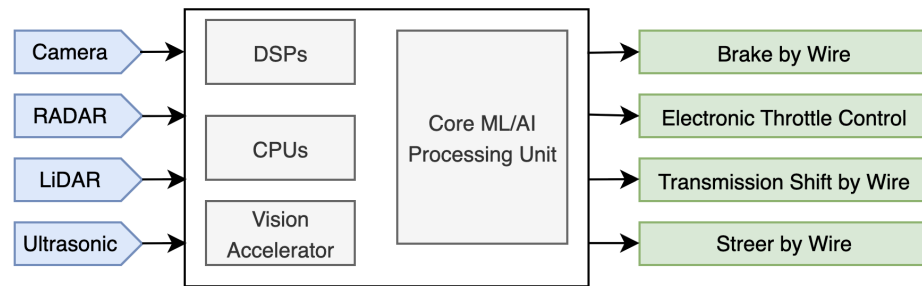


Figure 2. ADAS General Processing Structure [9].

This review provides a critical examination of the machine learning and knowledge extraction techniques that underpin modern autonomous driving systems. Rather than offering only a descriptive survey, we analyze the autonomous stack from a data-centric perspective, compare modular and end-to-end architectures, and examine how leading industry players translate these ideas into practical systems. We also identify the main technical barriers that still prevent reliable deployment, with particular attention to model robustness, interpretability, and rare long-tail scenarios [15,16].

The main contribution of this paper is a structured and critical synthesis of autonomous driving from the perspective of machine learning and knowledge extraction. Specifically, we: (i) organize the sensor and software stack according to the role of each component in the data-processing pipeline; (ii) compare classical modular pipelines with end-to-end learning frameworks; (iii) review the hardware and simulation tools used to train and validate these systems; and (iv) identify the open problems that still prevent reliable Level 4 and Level 5 autonomy, especially long-tail edge cases, adverse-weather robustness, and explainability.

The remainder of the paper is organized as follows. Section 2 reviews the sensor stack. Section 3 discusses computing platforms. Section 4 presents the artificial intelligence pipeline. Section 5 examines autonomous driving simulators. Section 6 examines industry case studies. Section 7 discusses challenges and future directions, and Section 8 concludes the paper.

While numerous surveys exist on autonomous driving, most focus either on perception algorithms [17–19], sensor technologies [20–22], or high-level system architectures in isolation [23–25]. In contrast, this review adopts a unified, data-centric perspective that jointly analyzes sensing modalities, AI architectures, system-level design choices, and industrial deployment strategies. Additionally, this work emphasizes the fundamental trade-offs between modular pipelines and end-to-end learning approaches, supported by recent advances such as transformer-based sensor fusion methods and simulation-based validation frameworks. This integrated perspective is particularly relevant given the current divergence in industry practices (e.g., vision-centric vs. multi-modal approaches), making a consolidated critical analysis both timely and necessary.

2. Review Methodology

This review follows a structured literature selection methodology to ensure coverage of the most relevant and recent advancements in autonomous driving. Publications were collected from major scientific databases, including IEEE Xplore, Scopus, ScienceDirect, and arXiv, focusing on works published between 2018 and 2026 to capture recent trends in deep learning and autonomous systems.

Keywords used during the search included “autonomous driving,” “sensors fusion,” “deep learning,” “transformers,” “end-to-end driving,” and “autonomous vehicle simulation.” Priority was given to highly cited papers, benchmark datasets, and industrial reports.

3. Sensor Stack

The perception system of an autonomous vehicle begins with data acquisition. No single sensor modality is sufficient for safe operation in complex driving environments, so modern autonomous systems rely on a complementary sensor stack in which each modality compensates for the weaknesses of the others [26]. This multi-modal design is essential for redundancy, robustness, and reliable perception in safety-critical driving tasks. The core data acquisition technologies include cameras, LiDAR, radar, and supporting localization sensors such as GNSS and IMU. Table 2 summarizes how sensor choice varies across common ADAS applications.

Table 2. Sensors used in ADAS and Their Applications.

ADAS Application	Primary Sensor(s)	Main Fucntion
Surround view monitoring	Camera (multiple wide-angle cameras)	Provides a 360° view around the vehicle
Park assistance	Camera, Ultrasonic	Detects nearby obstacles during parking maneuvers
Blind spot detection	Radar, LiDAR	Detects vehicles in adjacent lanes outside driver visibility
Rear collision warning	Radar, LiDAR	Monitors approaching vehicles from the rear
Rear cross-traffic alert	Radar, camers	Detects crossing vehicles when reversing
Automatic Emergency Braking (AEB)	Radar, LiDAR, Camera	Detects imminent collisions and automatically applied brakes
Pedestrian detection	Camera, LiDAR	Identifies pedestrians and vulnerable road users
Collision avoidance	Radar, LiDAR, Camera	Detects obstacles and assists in avoiding collisions
Traffic sign recognition	Camera	Recognizes traffic signs and speed limits
Adaptive cruise control (ACC)	Radar, LiDAR, Camera	Maintains safe distance from preceding vehicles
Lane departure warning (LDW)	Camera	Detects lane markings and warns drivers of unintended lane departure
Lane Keeping Assist (LKA)	Camera, Radar	Provides steering assistance to keep vehicles centered within lanes
Forward Collision Warning (FCW)	Radar, Camera	Warns drivers about potential frontal collisions
Driver Monitoring Systems (DMS)	Infrared Camera	Monitors driver attention and fatigue

3.1. Cameras: The Semantic Sensor

Cameras are the most ubiquitous perception sensors in autonomous driving because they are inexpensive and provide dense, high-resolution color and texture information that closely resembles human vision [27,28]. They are the primary source of semantic knowledge for machine learning models, especially for object classification, traffic light recognition, lane marking detection, and drivable-area segmentation. Table 3 lists commonly used camera models in the autonomous vehicle industry, while Table 4 summarizes representative stereo camera specifications.

The main limitation of camera-based perception is that it is inherently 2D. Depth estimation must be inferred rather than directly measured, which makes monocular perception difficult and computationally expensive [29,30]. In addition, camera performance degrades under poor lighting, glare, rain, and fog, creating a major robustness challenge for vision-based systems [31,32]. This is one reason the industry remains divided between camera-centric strategies and multi-modal designs. Tesla Vision is a prominent exam-

ple of the former, having shifted away from radar and LiDAR in favor of a vision-first architecture [33].

Table 3. Commonly used cameras in the autonomous vehicle industry.

Video Camera Name/Type	Optimal Environment	Field of View
Aspect 360 (Surround view camera)	Clear weather, parking, low-speed maneuvers	360°
Teledyne FLIR CMOS (RGB automotive camera)	Clear weather, daytime	80° × 64.4°
Lepton Thermal (LWIR)	−10 °C to +65 °C incl. fog and low light	Diagonal: 63.5° Horizontal: 50°
ZF's S-Cam4 (Tesla) Aptina AR0132/AR0136A/OmniVision OV10635	All conditions except dense fog	48°
Automotive HDR RGB Camera (e.g., ON Semi AR0231AT)	High dynamic range scenes, daylight to dusk	120°
Monocular Front-facing Camera	Highway and urban driving	30°–60°
Fisheye Camera	Close-range perception, surround view	180°–200°
Stereo Vision Camera (e.g., ZED, Mobileye)	Depth estimation, structured environments	90°
Near-Infrared (NIR) Camera	Night-time driving with IR illumination	40°–60°
Event-based Camera (Dynamic Vision Sensor)	High-speed motion, high-contrast scenes	~120°

Table 4. Representative stereo camera specifications

Model	Baseline (mm)	HFOV (°)	Range (m)	Resolution (MP)	FPS (Hz)
Intel RealSense D455 [34]	95	86	0.4–20	3.0	30–90
Intel RealSense D435 [35]	50	86	0.1–10	3.0	30–90
Carnegie MultiSense S21B [36]	210	68–115	0.4+	2.0–4.0	7.5–30
Carnegie Robotics KS21 [36]	210	135	0.5–20	2.3	30
Carnegie Robotics S27 [37]	270	91	1.4–30	2.3	30
Roboception RC Visard 160 [38]	160	61	0.5–3	1.2	0.8–25
Stereolabs ZED 2 [39]	120	110	0.3–20	4.0	15–100
Stereolabs ZED 2i [40]	120	110	0.2–20	4.0	15–100

3.2. LiDAR: The Geometric Sensor

Light Detection and Ranging (LiDAR) sensors provide precise geometric information by emitting laser pulses and measuring their return time to build a detailed 3D point cloud of the environment [9]. This makes LiDAR a cornerstone of many Level 4 systems because it supports high-precision mapping, localization, and 3D object detection. Table 5 summarizes commonly used LiDAR models and their specifications.

LiDAR has two major advantages. First, it provides direct distance measurements with high spatial accuracy. Second, as an active sensor, it is less dependent on ambient light and can operate in darkness. Its main limitations are cost, data volume, and sensitivity to adverse weather, especially fog and heavy rain. The high dimensionality of LiDAR point clouds also requires specialized neural architectures, such as PointNet- and voxel-based models, for efficient processing [41]. These constraints explain why LiDAR is often used together with cameras and radar rather than as a standalone modality.

Table 5. Commonly used LiDARs in the autonomous vehicle industry [42].

LiDAR	Accuracy	Range/Field of View	Power Consumption
Ouster OS-0	$\pm 1.5\text{--}5$ cm	50 m/90°	14–20 W
Luminar IRIS	1 cm	500 m/120°	15 W
Velodyne Alpha Prime	± 3 cm	245 m (best at 150 m)/360°	22 W
Teledyne CL-90	1 cm	176 to 600 m/64–90°	60 W
RoboSense RS-LiDAR-M1	± 5 cm	200 m/120°	18 W
Hesai Pandar40P	± 2 cm	200 m/360°	18 W
Livox Tele-15	± 2 cm	500 m/15°	12 W

3.3. Radar: The Robust Sensor

Radar is a cornerstone sensor technology in autonomous vehicles because it can estimate distance, velocity, and angle by emitting and receiving reflected radio waves. It is especially valuable in adverse weather and poor lighting, where optical sensors are less reliable. Automotive radar systems are commonly deployed in the 24 GHz, 60 GHz, 77 GHz, and 79 to 81 GHz bands [43]. Their use can be grouped into four main classes.

- Short-Range Radar (SRR): Used for parking assistance, blind spot detection, and collision avoidance at distances below 100 m with a wide field of view.
- Medium-Range Radar (MRR): Used for lane change assistance and cross-traffic alerts at intermediate distances of 100 to 200 m.
- Long-Range Radar (LRR): Used for adaptive cruise control and forward collision warning at ranges of up to 300 m.
- 4D Imaging Radar: Extends conventional radar by adding elevation information, which improves scene understanding and supports higher-resolution environmental modeling.

Radar offers three main benefits. It is robust in rain, fog, dust, and darkness. It provides direct velocity estimation through the Doppler effect. It is also generally less expensive than LiDAR. However, radar has lower angular resolution, limited object classification ability, and is susceptible to false positives and frequency interference in dense traffic. Because of these limitations, radar is best viewed as a complementary modality for motion estimation and redundancy rather than a standalone perception sensor. Table 6 shows some of the commercial radars and their corresponding power consumption.

Table 6. Representative automotive radar sensors and their characteristics

Model	Power (W)	Radar Type	Typical Application
Continental ARS 408-2	6.6	Long-range radar	Adaptive cruise control, forward collision warning
Bosch LRR3	4.0	Long-range radar	Highway driving, object tracking
Aptiv SRR2	6.0	Short-range radar	Blind-spot monitoring, lane change assistance
Aptiv MRR	4.5	Medium-range radar	Cross-traffic alert, surrounding vehicle detection
smartmicro UMRR-0A Type 29	3.7	Medium-range radar	Industrial and autonomous vehicle perception

3.4. Supporting Sensors for Localization

While cameras, LiDAR, and radar perceive the external environment, localization sensors estimate the vehicle's own state. GNSS and IMU are the most important supporting sensors in this category [44]. GNSS provides a global position reference, while the IMU measures acceleration and angular motion. Together, they support vehicle localization, especially when fused with other sensors to reduce drift and improve robustness in tunnels, urban canyons, and other GNSS-degraded settings. Table 7 lists example GNSS receivers, and Table 8 lists example ultrasonic sensors used in autonomous vehicles.

Table 7. Comparison of GNSS receivers for autonomous driving applications [45].

Model	Power (W)	Frequency	Positioning Mode	Typical Use Case
u-blox LEA-6T	0.5	L1	Standalone GNSS	Basic vehicle navigation
Trimble BD992	0.7	L1/L2	RTK GNSS	Precision positioning and surveying
NovAtel OEM729	0.9	Multi-band	GNSS + INS	Autonomous vehicle localization
AsteRx-m3 Pro+	1.8	Multi-band	Multi-constellation RTK	Lane-level positioning and HD mapping

Table 8. Representative ultrasonic sensors used in autonomous vehicles and their characteristics [45].

Model	Power (W)	Typical Application
Continental USR2-3P	0.5	Parking assistance and obstacle detection
NXP Semiconductors FXAS21002	0.7	Short-range environment sensing
STMicroelectronics VL6180X	1.0	Proximity sensing and object detection
TE Connectivity SENSONICS USI-60	2.0	Vehicle maneuvering and obstacle avoidance

Ultrasonic sensors are mainly used for short-range tasks such as parking and low-speed maneuvering. They are low-cost and useful for near-field detection, but they have limited range and are sensitive to interference. For this reason, they are usually treated as supporting sensors rather than core perception sensors.

3.5. Sensor Fusion and Industry Configurations

The central machine learning problem in the sensor stack is not only sensing but also fusion. Camera data provide semantics, LiDAR provides geometry, radar provides velocity and all-weather robustness, and GNSS/IMU provide vehicle state [46]. The challenge is to combine these heterogeneous streams into a single representation that is accurate, low-latency, and fault-tolerant. Fusion can be implemented at the raw data level, feature level, or decision level, and each strategy involves different trade-offs in complexity, latency, and interpretability [22,47].

As highlighted by Matos et al. [48], the reliability of autonomous driving systems is closely linked to the robustness of their sensing components. Ultrasonic sensors are susceptible to interference-induced range estimation errors, whereas radar systems may produce false detections due to signal reflections and mutual interference among radars operating at similar frequencies. LiDAR sensors can experience perception degradation in adverse weather conditions and inaccuracies caused by highly reflective surfaces. Camera sensors are particularly affected by lighting variability and image degradation resulting from rain, fog, or snow, which can reduce object detection accuracy and lead to misinterpretation by perception algorithms. GNSS-based localization systems face challenges including timing errors, signal jamming and spoofing, multipath effects, and satellite orbit uncertainties, all of which contribute to localization errors. Furthermore, IMUs inherently suffer from drift and cumulative error growth, reducing localization accuracy during prolonged operation. These limitations highlight the importance of sensor fusion techniques to improve the reliability and robustness of autonomous vehicle perception and localization systems [48].

Leading companies adopt different sensor philosophies, reflecting an ongoing debate about the best data acquisition strategy for autonomous driving. These configurations show the trade-off between cost, richness of sensing, and redundancy, see Table 9 for more details.

Overall, the sensor stack should be viewed as the data foundation of autonomous driving rather than as a list of hardware components. Its value lies in how effectively the vehicle converts raw sensory input into reliable information for perception, prediction, and planning.

Table 9. A comparative overview of sensor suite configurations from leading autonomous vehicle developers.

Company/Platform	Camera Count	LiDAR Count	Radar Count	Ultrasonic Count	Primary Philosophy
Waymo (6th Gen)	13	4	6	Yes	Multi-modal fusion for maximum redundancy
Tesla (Hardware 4)	8	0	1	12	Vision-centric, end-to-end learning
Baidu (Apollo RT6)	12	8	–	–	Dense LiDAR and camera suite for robotaxis
WeRide (Sensor Suite 5.0)	12	7 (solid-state)	–	–	Modular, lightweight design for various vehicles
Pony.ai (PonyAlpha X)	7	4	4	0	Multi-modal fusion with strong LiDAR and radar focus

3.6. Critical Comparison of Sensor Modalities

Although each sensor modality offers distinct advantages, their integration reflects fundamental trade-offs rather than complementary strengths alone. Cameras provide rich semantic understanding but struggle with depth estimation and adverse weather [49]. LiDAR delivers precise geometric information but at high cost and with sensitivity to environmental conditions such as fog and rain. RADAR offers robustness and velocity estimation but lacks sufficient resolution for fine-grained object classification.

A key tension in current system design lies between cost-efficiency and redundancy. Vision-only systems, such as Tesla’s approach, reduce hardware cost and complexity but place significant reliance on learned depth estimation and large-scale data [50]. In contrast, multi-modal systems used by Waymo and Baidu emphasize safety through redundancy but incur higher computational and economic costs.

Recent research trends suggest a hybrid trajectory, where multi-modal systems are optimized using learned fusion techniques. For example, transformer-based fusion models enable more efficient integration of heterogeneous sensor inputs, potentially reducing reliance on expensive sensors while maintaining robustness [51].

3.7. Sensory System Summary

This section demonstrated that autonomous vehicle perception relies on the complementary strengths of multiple sensing modalities rather than on any single sensor technology. Cameras provide rich semantic understanding, LiDAR offers precise geometric information, radar contributes robust velocity estimation and all-weather operation, while GNSS/IMU systems support localization. The analysis reveals a fundamental trade-off between sensing richness, redundancy, and cost. Vision-centric approaches offer lower hardware complexity but rely heavily on learning-based inference, whereas multi-modal systems provide stronger redundancy and robustness at increased operational and computational cost. Consequently, the future of autonomous driving is likely to depend not on selecting a single dominant sensor but on developing more efficient sensor fusion architectures that maximize information gain while minimizing cost and complexity.

4. Computing Platforms

High-performance, low-latency processing is the backbone of every autonomous driving system. The perception, prediction, and planning pipeline must ingest multi-modal sensor streams, execute deep-learning inference, and produce deterministic actuation commands within a tight real-time budget. To meet these demands, modern autonomous

vehicles rely on heterogeneous computing platforms that combine general-purpose processors with specialized accelerators [6,9].

4.1. General-Purpose CPUs and GPUs

Central Processing Units (CPUs) provide the flexibility required for system orchestration, sensor-data pre-processing, and safety-critical control loops. Multi-core ARM Cortex-A family or x86-64 cores are common in production vehicles because they support rich operating systems such as Linux and QNX and allow rapid software updates [6]. However, CPUs alone are not efficient enough for the computational burden of modern convolutional neural networks, especially when perception workloads must operate in real time.

For this reason, Graphics Processing Units (GPUs) are commonly integrated to accelerate deep-learning inference. Their large degree of parallelism makes them well suited to image processing, feature extraction, and tensor-heavy workloads. High-end automotive GPUs, such as NVIDIA Ampere-based devices used in the DRIVE Orin family, include thousands of CUDA cores and Tensor Cores that support mixed-precision computation and high-throughput matrix operations [9]. In practice, CPUs and GPUs are often used together, with the CPU handling system-level tasks and the GPU handling the bulk of neural inference.

4.2. Domain-Specific ASICs and Accelerators

Although GPUs offer strong flexibility, the demand for lower latency and better energy efficiency has driven the adoption of domain-specific accelerators. Field-Programmable Gate Arrays (FPGAs) provide a semi-customizable solution that allows engineers to tailor data paths, precision formats, and parallelism to specific neural network models. This makes FPGAs attractive for quantized or lightweight deep networks, where custom hardware mapping can significantly improve efficiency [52].

Application-Specific Integrated Circuits (ASICs) represent the most specialized end of the spectrum. These chips are designed for narrow classes of workloads, often focusing on matrix and vector operations used in deep learning. Their key advantage is that they deliver high performance at low power consumption, but this comes at the cost of flexibility. Once fabricated, an ASIC cannot be easily repurposed for new models or new workloads, which makes it best suited to stable, high-volume deployment scenarios [9,53].

In autonomous vehicles, ASICs are especially important for dedicated vision pipelines, neural-network inference, and sensor-specific pre-processing. They are often used when the objective is to maximize throughput while staying within strict thermal and power budgets.

5. Artificial Intelligence Pipeline

Artificial intelligence is the software core of modern autonomous driving systems. Machine learning models extract patterns from sensor data, deep learning models perform perception on complex unstructured inputs, and generative methods can support synthetic data creation and scenario augmentation. In practice, the autonomy stack is usually organized as a modular pipeline in which perception, prediction, and motion planning are executed sequentially, with each stage contributing a distinct representation of the driving scene.

As shown in Figure 3, the classical AI pipeline transforms raw multi-modal sensor inputs into safe vehicle actions. The pipeline first constructs an environmental model, then forecasts the future behavior of other agents, and finally computes a trajectory for the ego-vehicle. This structure remains widely used because it is interpretable and allows each module to be tested separately.

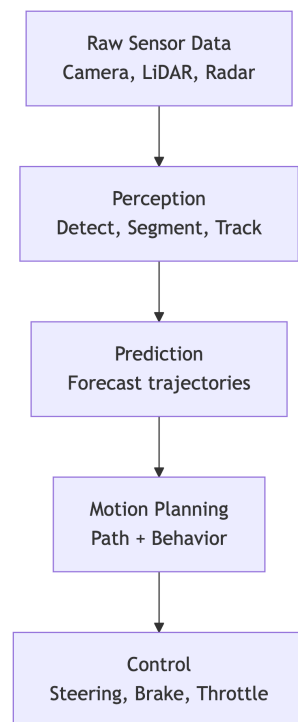


Figure 3. High-level block diagram of the classical sequential AI pipeline for autonomous vehicles.

5.1. Perception

Perception is the foundational layer of the AI pipeline. It processes raw data from cameras, LiDAR, and radar to build a structured, machine-readable representation of the surrounding world [54]. Because errors at this stage propagate to all downstream modules, perception quality has a direct impact on the safety of the complete system. Deep learning, especially convolutional neural networks, has largely replaced classical computer vision methods in this domain [2].

The main perception tasks are object detection and semantic segmentation. Object detection identifies and localizes relevant road users and infrastructure, including vehicles, pedestrians, cyclists, traffic lights, and signs. Semantic segmentation provides a denser understanding of the scene by assigning labels to every pixel or point, which is essential for identifying the drivable area and understanding the scene context [55]. State-of-the-art systems increasingly rely on multi-sensor fusion to improve robustness against the failure modes of individual modalities.

The large volume of sensor data also makes perception a hardware challenge because these models must run in real time on embedded automotive platforms.

5.2. Prediction and Motion Planning

Once perception established what is in the environment and where it is located, the next stages must determine what those objects are likely to do and how the ego-vehicle should respond [56]. Prediction is therefore responsible for forecasting the future motion of dynamic road users such as vehicles and pedestrians. Early systems often relied on simpler temporal models, while modern approaches use learned predictors that estimate multiple likely future trajectories for each actor, together with their associated probabilities.

Motion planning uses the predicted scene evolution and the current environmental model to compute a safe, legal, and comfortable path for the vehicle [57]. This problem can be divided into global planning, which determines the overall route, and local planning, which selects immediate tactical actions based on the current scene [58,59]. In most systems,

local planning is formulated as an optimization problem that balances collision avoidance, traffic-law compliance, and passenger comfort. Search-based approaches, such as state lattices, further discretize the problem into a graph-search formulation [60].

The output of this stage is a trajectory that can be executed by the vehicle control system through steering, acceleration, and braking commands.

5.3. End-to-End Learning

The classical modular pipeline is contrasted with the end-to-end learning paradigm. In this approach, a single neural network maps raw sensor inputs directly to vehicle control outputs, such as steering angle and acceleration, thereby collapsing perception, prediction, and planning into one learned model [33]. This approach is most closely associated with Tesla's Full Self-Driving philosophy, which relies on large-scale fleet data and imitation learning to learn driving behavior from examples.

Recent advances in end-to-end autonomous driving have been significantly influenced by transformer-based architectures that improve multi-modal reasoning. A representative example is TransFuser [61], which introduces a transformer-based sensor fusion mechanism that integrates camera and LiDAR features using attention across multiple spatial resolutions. Unlike traditional geometric fusion approaches that rely on local feature projections, TransFuser captures global context and interactions between distant objects, which is critical in complex urban scenarios with dense traffic.

Experimental results demonstrate that such attention-based fusion mechanisms substantially reduce collision rates by up to 48% compared to prior fusion approaches while improving overall driving performance. This highlights a key evolution in autonomous driving models: the transition from convolution-based feature extraction to attention-based architectures capable of modeling global scene relationships.

The main attraction of end-to-end learning is its ability to capture complex driving behavior that may be difficult to hand-engineer into a modular system. However, the approach also introduces major limitations that are central to the current debate in autonomous driving. Large neural networks are difficult to interpret and debug, which complicates safety validation. Their performance can also degrade in rare long-tail scenarios that are not adequately represented in the training data. In addition, these models may learn spurious correlations instead of true causal relationships, which creates serious risks when they encounter unfamiliar situations.

For these reasons, the industry remains divided between modular architectures, which emphasize explicit safety constraints and interpretability, and end-to-end or hybrid approaches, which emphasize data-driven generalization.

Table 10 highlights a fundamental trade-off: modular pipelines prioritize interpretability and safety assurance, while end-to-end systems emphasize scalability and data-driven generalization. Hybrid architectures that combine these strengths are increasingly viewed as the most promising direction.

Table 10. Comparison of Modular vs. End-to-End Architectures.

Criterion	Modular Pipeline	End-to-End Learning
Interpretability	High	Low
Data Requirements	Moderate	Very High
Robustness	High (with redundancy)	Sensitive to data bias
Safety Validation	Easier	Difficult
Scalability	Moderate	High
Error Isolation	Possible	Not Possible

5.4. Datasets for Autonomous Driving

Publicly available datasets are essential for training and evaluating autonomous driving systems, and the most widely used ones are summarized in Table 11 [62]. These datasets cover perception, mapping, prediction, and planning tasks, and they differ in sensor setup, geography, and traffic conditions. They are indispensable for benchmarking model performance and improving generalization across urban, rural, and highway environments.

Table 11. Publicly available datasets for autonomous driving research [62].

Dataset	Problem Space	Sensor Set Up	Location	Traffic Condition
NuScenes	3D object detection, tracking, online vectorized map creation	Camera, radar, lidar, GPS, IMU	Boston, Singapore	Urban
KITTI	3D object detection, tracking, SLAM	Camera, lidar, GPS, IMU	Karlsruhe, Germany	Urban, Rural
Udacity	3D object detection, tracking	Camera, lidar, GPS, IMU	Mountain View, USA	Rural, Urban
Cityscapes	Semantic segmentation	Camera, lidar, GPS, IMU	Switzerland, France	Urban
Ford	3D object detection, tracking	Camera, lidar, GPS, IMU	Michigan	Urban
Daimler pedestrian	Pedestrian detection, classification, segmentation, path prediction	Mono and stereo camera	Europe, China	Urban
BDD	2D/3D object detection, tracking, semantic segmentation	Camera	USA	Urban, Rural
Oxford	3D tracking, 3D object detection	Camera, lidar, GPS, IMU	Oxford	Urban, Highway

The datasets used in the field are typically grouped according to the tasks they support, including perception, mapping, prediction, and planning, as shown in Figure 4 [63]. Their diversity is useful, but it also means that many benchmarks remain biased toward well-structured driving scenes, leaving rare edge cases and difficult weather conditions underrepresented.

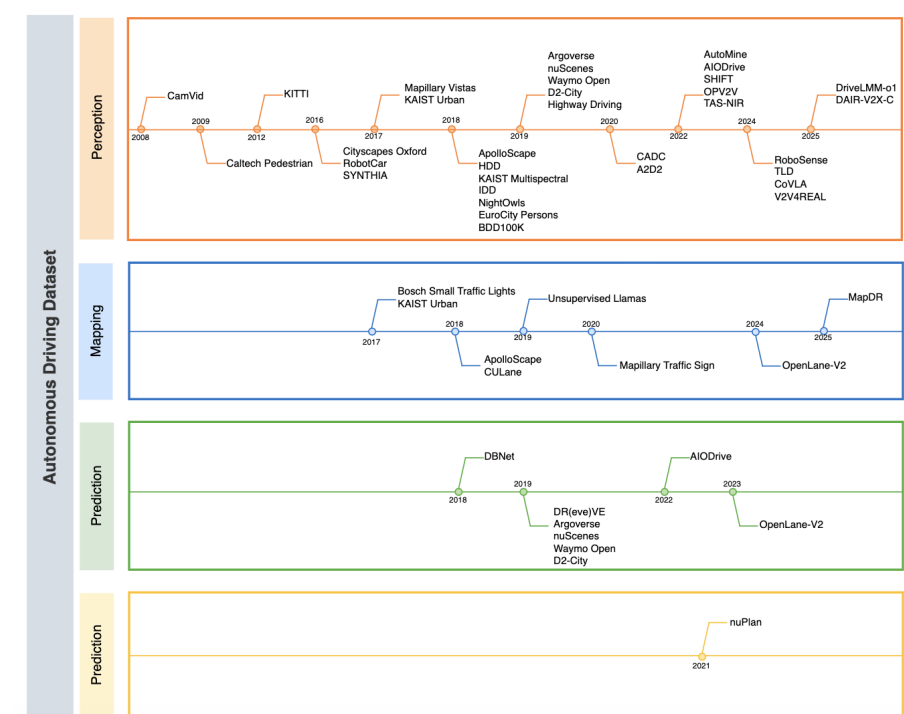


Figure 4. Timeline of dataset on autonomous driving.

Figure 5 shows the distribution of sensor modalities across commonly used public autonomous-driving benchmark datasets. Each dataset is counted once for each sensor modality it officially provides. The statistics reflect trends within widely adopted benchmarks rather than an exhaustive census of all released datasets. The datasets included are KITTI, Cityscapes, BDD100K, ApolloScape, Mapillary Vistas, nuScenes, Waymo Open Dataset, Argoverse, Argoverse2, PandaSet, SemantickITTI, KAIST Multispectral, FLIR ADAS, DSEC, and MVSEC.

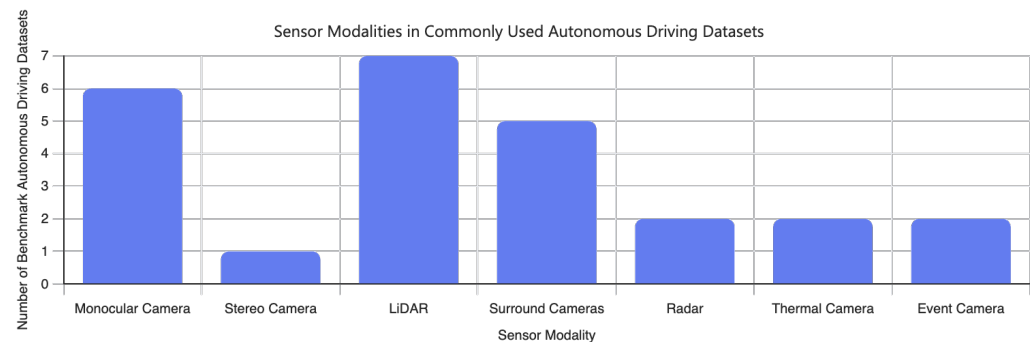


Figure 5. Distribution of sensor modalities across commonly used public autonomous driving benchmark datasets.

5.5. Evolution of AI Architectures

The evolution of AI models in autonomous driving reflects a shift from task-specific architectures toward unified and data-driven frameworks. Early systems relied heavily on convolutional neural networks (CNNs) for perception tasks such as object detection and segmentation. These were later extended with recurrent architectures, including LSTMs, to capture temporal dependencies for trajectory prediction.

More recently, transformer-based architectures have emerged as a dominant paradigm due to their ability to model long-range dependencies and multi-modal interactions. Compared with CNNs, transformers provide greater representational flexibility but at a higher computational cost. This trade-off has driven research into efficient attention mechanisms and hybrid architectures.

Finally, end-to-end learning systems represent a further abstraction, integrating perception, prediction, and planning into a single model. While these systems offer scalability and improved generalization, they remain challenging in terms of interpretability and safety validation.

5.6. Architectures Summary

This section examined the evolution of artificial intelligence architectures for autonomous driving, from modular perception—prediction—planning pipelines to modern end-to-end learning systems. While modular approaches remain advantageous for interpretability, debugging, and safety validation, end-to-end methods offer increased scalability and the ability to learn complex behaviors directly from large-scale datasets. Recent transformer-based architectures further enhance multi-modal reasoning by capturing long-range contextual dependencies that were difficult to model using traditional convolutional networks. However, no single paradigm currently satisfies all requirements of safety, robustness, explainability, and generalization. The current research landscape therefore points toward hybrid architectures that combine the transparency of modular systems with the learning capacity of data-driven approaches.

6. Autonomous Driving Simulators

Simulation platforms are essential tools in autonomous vehicle development because they enable repeatable, scalable, and safe testing in virtual environments that would be impractical or unsafe to reproduce on public roads. They are used for perception validation, planning evaluation, sensor testing, and scenario generation, and they play a central role in bridging the gap between offline development and real-world deployment. In practice, the simulator landscape is divided into open-source platforms, which are widely used in academic research, and proprietary commercial platforms, which are often used for industrial-scale validation.

6.1. Open-Source Simulators

6.1.1. AirSim

AirSim, developed by Microsoft, uses the Unreal Engine to provide high-fidelity visual and physical simulation for autonomous driving and robotics research [64]. Its modular architecture supports customizable environments, vehicle dynamics, sensor models, and hardware-in-the-loop testing, which makes it useful for perception and control experiments.

6.1.2. CARLA (Car Learning to Act)

CARLA is an open-source simulator jointly developed by Intel Labs and the Computer Vision Center for the development, training, and validation of autonomous systems [65]. Built on Unreal Engine 4, it supports photorealistic urban scenes, configurable traffic, pedestrians, weather conditions, and a broad sensor suite, making it one of the most widely used benchmarks for autonomous driving research.

6.1.3. Baidu Apollo Simulator

The Apollo ecosystem includes an integrated simulator for testing autonomous driving systems within its platform [66]. It provides open interfaces for vehicle hardware, sensors, and computing specifications, and a cloud-based data pipeline that supports large-scale model training and virtual validation before deployment.

6.1.4. Autoware

Autoware is an open-source autonomous driving software stack built on the Robot Operating System (ROS) [67]. Its modular architecture includes dedicated components for perception, localization, planning, and control, and it is designed to support integration with real vehicle hardware for both simulation and on-road testing.

6.1.5. Gazebo

Gazebo is a general-purpose 3D robotics simulator used for testing autonomous vehicles and other robotic systems. It supports multiple physics engines, such as DART, ODE, and Bullet, and includes models for sensors such as cameras and LiDAR, which makes it useful for robotics-oriented autonomy research [68].

6.2. Proprietary and Commercial Simulators

6.2.1. CarCraft

CarCraft is Waymo's proprietary simulation platform and is used exclusively for training and validating autonomous driving software. It is designed for large-scale scenario generation and closed-loop testing, allowing Waymo to accumulate extensive virtual driving experience by continuously creating and modifying traffic situations [69,70].

6.2.2. Ansys Autonomy

Ansys Autonomy supports real-time closed-loop simulation with multiple sensors, traffic objects, and dynamic environments. It provides physically accurate 3D scene modeling and can import high-precision maps such as OpenStreetMap to generate road networks for testing [71].

6.2.3. Cognata

Cognata is a cloud-based simulation platform that uses digital-twin concepts and deep learning to generate high-fidelity driving scenarios. It is designed to accelerate autonomous vehicle commercialization by enabling large-scale validation across diverse conditions, with a customizable library of scenes that users can modify [72].

6.2.4. MATLAB/Simulink

MATLAB v R2024a and Simulink provide the Automated Driving Toolbox, which offers tools for perception development, sensor fusion, high-definition map access, and ADAS prototyping [63]. Their strength lies in their broader engineering ecosystem, making them useful for algorithm development, validation, and rapid integration with control and signal-processing workflows.

6.3. Critical Perspective

Overall, simulation platforms are indispensable because they support safe testing, repeatability, and long-tail scenario generation. Open-source tools are especially valuable for research flexibility and benchmarking, while proprietary platforms are stronger in terms of scale, fidelity, and industrial validation. The main limitation across all simulators is the sim-to-real gap, meaning that performance in virtual environments does not always transfer directly to real-world driving conditions [63,69].

Recent work has also highlighted limitations in existing simulation environments, particularly their tendency to assume idealized perception and control conditions. For instance, the SimLKAS framework introduces a more realistic simulation-based evaluation approach for lane-keeping systems by incorporating perception uncertainties, lateral deviations, and environmental variability.

Unlike traditional simulators such as CARLA, which often assume perfect lane detection and near-ideal vehicle trajectories, SimLKAS models real-world imperfections, including noisy sensor inputs and driver-like trajectory variations. It also enables the systematic generation of diverse scenarios (e.g., rain, fog, and nighttime conditions), allowing quantitative assessment of performance degradation under different factors. This demonstrates an important trend: simulation is evolving from simple testing environments toward data-driven validation platforms that explicitly model uncertainty and system failure modes, which is essential for safety-critical applications.

Simulation has become an indispensable component of autonomous driving development by enabling safe, repeatable, and large-scale evaluation of complex driving scenarios. Open-source simulators provide flexibility and accessibility for research, whereas commercial platforms offer higher fidelity and industrial-scale validation capabilities. Despite their importance, current simulation frameworks remain limited by the sim-to-real gap, whereby successful performance in virtual environments does not necessarily guarantee equivalent behavior in real-world conditions. Recent trends toward uncertainty-aware and validation-focused simulation frameworks indicate a growing recognition that realistic safety assessment requires not only scenario diversity but also accurate modeling of perception errors, environmental variability, and rare edge cases. Consequently, simulation

should be viewed as a complement to, rather than a replacement for, real-world testing and validation.

7. Industry Landscape and Leading Approaches

The autonomous vehicle industry is not advancing along a single technological path. Instead, leading companies are pursuing different combinations of sensor configuration, software architecture, and deployment strategy, creating a real-world comparison between sensor-rich modular systems and vision-centric end-to-end approaches [73,74]. These differences are not only technical choices but also reflections of each company's safety philosophy, target operational design domain, and commercialization strategy.

7.1. Waymo: The Multi-Modal Redundancy Approach

Waymo is one of the clearest examples of a safety-first, multi-modal strategy. Its platform is built around sensor redundancy, modular perception, and explicit planning and control layers, which together support interpretable and verifiable autonomy [75,76]. In this architecture, perception estimates the surrounding scene, prediction forecasts the behavior of other agents, and planning computes a safe trajectory under explicit constraints.

Perception in Waymo-style systems relies on deep learning models that process camera, LiDAR, and radar data to detect and track vehicles, pedestrians, traffic lights, and lane structure in real time. Prediction modules then estimate future trajectories by combining motion history with scene context, while planning and control apply rule-based logic and optimization to produce safe, comfortable vehicle motion. The strength of this approach is that each stage can be validated independently, which improves interpretability and safety assurance.

7.2. Tesla: The Vision-Centric, End-to-End Strategy

Tesla follows a different philosophy centered on camera-only perception and end-to-end learning. The core idea is to use large-scale fleet data and neural networks to map raw visual input directly to driving decisions, rather than relying on a fully modular perception-prediction-planning stack [33,74]. This strategy prioritizes scalability and data-driven generalization, but it also places greater pressure on the vision system to handle adverse weather, poor lighting, and rare edge cases.

Tesla's hardware strategy still reflects a strong emphasis on compute redundancy and real-time inference. The FSD computer integrates multiple processing elements, including CPUs, GPUs, neural-network accelerators, and camera interfaces, to support sensor preprocessing and post-processing. However, the important architectural point for this section is not the exact chip specification but the fact that Tesla couples a vision-centric sensing philosophy with tightly integrated onboard computing. That combination is what differentiates it from multi-modal modular competitors.

7.3. Leaders in China's Rapid Deployment

China has emerged as a major autonomous driving market, with companies such as Baidu, WeRide, and Pony.ai advancing rapidly under supportive regulation and strong commercialization pressure [77]. These firms generally retain multi-modal sensing, but they differ in how much emphasis they place on LiDAR density, fusion strategy, and platform integration.

7.3.1. Baidu Apollo

Baidu Apollo is a leading robotaxi platform in China and has deployed Apollo Go in several major cities [78]. Its vehicles use multi-modal sensor suites that combine cameras, radar, and multiple LiDAR units to support robust 3D mapping, localization, and object

detection. Compared with Tesla’s vision-centric model, Baidu’s approach remains closer to the redundancy-oriented philosophy used by Waymo.

7.3.2. WeRide

WeRide combines multi-sensor fusion with high-definition maps and a proprietary middleware layer to support scalable autonomy [41]. Its system is designed for reliable localization in challenging urban settings such as tunnels, bridges, and dense city streets. A notable feature of its approach is the use of dual fusion pipelines, one vision-centric and one LiDAR-centric, which improves resilience when one modality is degraded.

7.3.3. Pony.ai

Pony.ai also follows a multi-modal strategy, combining cameras, LiDAR, and radar with heterogeneous onboard computing to support both performance and redundancy [79,80]. Its architecture reflects a pragmatic production strategy: use multiple sensing modalities, pair them with dedicated computing resources, and maintain enough flexibility to operate across different vehicle platforms and operating environments.

7.4. Operational Design Domain Comparison

These companies also differ in their operational design domains, which helps explain why their technology choices diverge. Waymo is strongest in carefully mapped and highly controlled urban environments, Tesla targets broad consumer deployment with driver-supervised automation, and the Chinese robotaxi companies often focus on geofenced commercial deployment in selected cities, Table 12 shows the characteristics for selected companies’ AVs.

Table 12. ODD characteristics for selected autonomous vehicle companies [81].

Vehicle Company	Country	Environment	Operational Conditions	Driving Scenarios
Waymo Driver	United States	Sunny, light rain/snow	Moderate traffic density	Lane changes, highway merging/exiting, multi-lane highways, rural roads, daytime/nighttime, dynamic route planning
Tesla Autopilot	United States	Clear weather	Limited traffic density, specific speed ranges	Lane markings, driver supervision required
Baidu Apollo	China	Clear weather, limited traffic density	Daytime, nighttime	Highways and city streets in specific zones, lane changes, highway merging/exiting, traffic light and stop sign recognition, intersection navigation, low-speed maneuvering
WeRide	China	Clear weather	Daytime, nighttime	Limited-access highways and urban streets, lane changes, highway merging/exiting, traffic light and stop sign recognition, intersection navigation, automated pick-up/drop-off
Pony.ai	China	Diverse weather, including heavy rain/snow	High traffic density, frequent stops and turns	Narrow city streets, residential areas, parking lots, low speeds, geo-fenced zones, pedestrian and cyclist detection

Overall, the industry does not yet agree on a single best architecture for autonomy. The current landscape suggests three competing paths: redundant multi-modal sensing, vision-centric end-to-end learning, and hybrid systems that combine the interpretability of modular pipelines with the scale of neural learning.

8. Challenges and Outlook

Despite remarkable progress, the path to reliable Level 4 and Level 5 autonomy remains blocked by major technical, legal, and societal challenges. Overcoming these barriers is essential if autonomous vehicles are to deliver their promised gains in safety, efficiency, and mobility.

8.1. Safety Validation and Edge-Case Handling

The foremost challenge is demonstrating that an autonomous system is safer than a human driver. This requires more than functional competence because safety claims must be supported by statistically meaningful evidence. Since rare crashes are difficult to observe at the scale required for on-road validation, companies rely heavily on simulation and large-scale virtual testing to evaluate performance under controlled but diverse conditions [82,83].

However, simulation cannot eliminate the reality gap. Virtual environments may approximate physics and traffic behavior, but they cannot fully reproduce the complexity of real-world driving [84]. This limitation becomes especially important for edge cases, which are rare, unexpected, and often unusual scenarios that fall outside the training distribution. These cases may involve unusual debris, ambiguous road geometry, or unpredictable human behavior. Because it is impossible to enumerate every possible driving situation, robust autonomy depends on generalization, diverse data collection, and scenario generation rather than exhaustive rule coverage [85].

Interpretability is another major obstacle. End-to-end deep learning models can be powerful, but their internal decision-making is often difficult to inspect. This black-box behavior complicates debugging, certification, and regulatory approval because developers and authorities need to understand why a system produced a specific action [86].

8.2. Cybersecurity and Privacy

As connected cyber-physical systems, autonomous vehicles are exposed to security threats that go beyond traditional automotive risks.

- **Cybersecurity Threats:**
Autonomous vehicles can be targeted through sensor spoofing, communication hijacking, software exploitation, and adversarial attacks against perception models. A defense-in-depth strategy is therefore required, including secure coding, encryption, penetration testing, redundancy, and continuous monitoring to detect and respond to malicious behavior.
- **Data Privacy:** Autonomous vehicles collect large volumes of sensitive data, including high-definition camera imagery, precise location history, and potentially in-cabin information. This creates important privacy concerns, especially when such data are used for training, validation, or fleet learning. Clear rules for data collection, storage, and access are necessary to maintain public trust.

8.3. Regulatory Gaps and Public Acceptance

The technical progress of autonomous driving has outpaced the legal and social frameworks needed to govern it. A major challenge is the lack of harmonized regulations for testing, deployment, and certification. In the United States, regulation remains fragmented across states, which creates a complex compliance landscape for companies operating at a

national scale. Liability is also unresolved: if a crash occurs, responsibility may rest with the owner, the manufacturer, the software developer, or some combination of the three.

Public acceptance remains equally important. Many users still express concern about system failure, cybersecurity, and the loss of human control. Building trust will require transparent reporting, clearly stated operational limits, and repeated demonstrations of real-world reliability.

8.4. Formal Safety Validation Frameworks

A major limitation in current autonomous driving systems is the absence of standardized, scalable safety validation frameworks. While traditional automotive safety relies on standards such as ISO 26262 [87], these were not designed for data-driven AI systems.

Recent research emphasizes the need for simulation-based verification and validation (V&V) frameworks that can systematically evaluate system behavior across diverse operational scenarios. For example, simulation-based frameworks such as SimLKAS enable controlled experimentation under rare and extreme conditions that are difficult to capture in real-world testing.

This shift toward scenario-based testing and data-driven validation reflects a broader trend in safety-critical AI systems: moving from empirical validation to statistically grounded safety guarantees.

8.5. Future Trends

Addressing these challenges will require continued innovation across models, data, and infrastructure.

- **Advanced AI Paradigms:** Researchers are exploring foundation models for driving and modular end-to-end planning frameworks. These approaches aim to improve generalization while preserving some of the interpretability and safety benefits of modular systems.
- **Data Engines and Continuous Learning:** Leading companies are building data engines that create a feedback loop between real-world operation, scenario mining, and model retraining. This allows systems to improve over time by focusing on the most informative and difficult cases.
- **Infrastructure and Connectivity:** Vehicle-to-Everything (V2X) communication and smart infrastructure can extend awareness beyond onboard sensors. By exchanging information with other vehicles and traffic infrastructure, autonomous systems can improve coordination, anticipate hazards earlier, and increase traffic efficiency.

Overall, the central challenge facing autonomous driving is not the absence of sensing, computing, or learning capabilities but the integration of these components into systems that can be trusted under the vast diversity of real-world conditions. The comparison of current industrial and academic approaches suggests that future progress will depend less on incremental improvements in individual algorithms and more on achieving robust generalization, explainable decision-making, scalable validation, and regulatory acceptance. These challenges collectively make autonomous driving one of the most demanding safety-critical applications of artificial intelligence.

9. Conclusions

This review examined the artificial intelligence and machine learning foundations of modern autonomous driving systems, with emphasis on the sensor stack, computing platforms, AI pipeline, simulation tools, industry strategies, and open challenges. Across the stack, the central theme is clear: autonomous driving is not a single-model problem

but a multi-layered systems problem that requires robust sensing, efficient computation, reliable prediction, and safe decision-making.

Our analysis showed that no single sensor modality is sufficient on its own. Cameras, LiDAR, radar, GNSS, IMU, and ultrasonic sensors each contribute complementary information, making sensor fusion a core requirement for reliable perception. At the same time, industry practice is still split between modular, multi-modal systems such as Waymo's and vision-centric, end-to-end strategies such as Tesla's, while companies like Baidu, WeRide, and Pony.ai pursue hybrid approaches that balance redundancy with scalability.

The classical AI pipeline remains important because modular perception, prediction, and planning stages are easier to interpret, validate, and debug. End-to-end learning offers a powerful alternative, especially when large-scale fleet data are available, but it also introduces significant concerns related to interpretability, causal ambiguity, and rare edge cases. As a result, the current state of the field suggests that neither paradigm is universally superior, and future progress will likely come from hybrid systems that combine the strengths of both.

On the hardware side, autonomous vehicles are increasingly built on heterogeneous systems-on-chip that integrate CPUs, GPUs, and domain-specific accelerators to satisfy real-time and energy-efficiency constraints. Simulation platforms such as CARLA, AirSim, Gazebo, CarCraft, Cognata, and MATLAB/Simulink have also become indispensable for training, validation, and scenario generation, but they cannot fully eliminate the sim-to-real gap. For this reason, real-world deployment still depends on careful validation across both virtual and physical environments.

Several major challenges remain unresolved. Safety validation for Level 4 and Level 5 autonomy requires stronger statistical guarantees than current road testing can provide. Cybersecurity, privacy, regulation, and public acceptance also remain critical barriers to large-scale deployment. These issues make clear that autonomous driving is not only a technical challenge but also a legal, ethical, and societal one.

Looking ahead, foundation models, modular end-to-end planning, continuous learning from fleet data, and vehicle-to-everything communication are likely to shape the next phase of development. The most successful autonomous systems will be those that combine the performance of deep learning with the interpretability, redundancy, and safety assurance of modular design. Progress toward full autonomy will depend on coordinated advances in algorithms, hardware, simulation, regulation, and trust.

Author Contributions: E.K.: Conceptualization; Methodology; Investigation; Data Curation; and Writing—Original Draft Preparation. F.F.: Methodology; Investigation; Writing—Original Draft Preparation; and Visualization. A.A.: Resources. O.S.: Conceptualization; Writing—Review and Editing; Supervision; and Project Administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No data were used.

Acknowledgments: The authors acknowledge the use of generative AI tools (Microsoft Copilot) for minor language polishing and improving the clarity of selected paragraphs. The AI tool was used solely for linguistic refinement. All scientific content, analysis, and conclusions were developed, verified, and approved entirely by the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Brunner, P.; Vogl, S. Extracting Product Improvement Insights from Social Media Comments Using Machine Learning: A Case Study in the Automotive Industry. *Mach. Learn. Knowl. Extr.* **2026**, *8*, 42.
2. Wang, X.; Maleki, M.A.; Azhar, M.W.; Trancoso, P. Moving forward: A review of autonomous driving software and hardware systems. *arXiv* **2024**, arXiv:2411.10291.
3. Studer, S.; Bui, T.B.; Drescher, C.; Hanuschkin, A.; Winkler, L.; Peters, S.; Müller, K.R. Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 392–413. <https://doi.org/10.3390/make3020020>.
4. Organization, W.H. *Global Status Report on Road Safety 2018*; World Health Organization: Geneva, Switzerland, 2019.
5. Shalash, O.; Khatab, E.; El-Agamy, A.; Elmokadem, L.; Abouelsaad, Y.; Zaki, J.; El-Sayed, M.; and Said, H. Benchmarking YOLOv8–YOLOv12 for Real-Time Object Detection on Single-Board Computers. *Preprints*, 2026.
6. Liu, L.; Lu, S.; Zhong, R.; Wu, B.; Yao, Y.; Zhang, Q.; Shi, W. Computing systems for autonomous driving: State of the art and challenges. *IEEE Internet Things J.* **2020**, *8*, 6469–6486.
7. SAE, T. *Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems-j3016*; Society of Automotive Engineers: On-Road Automated Vehicle Standards Committee: Warrendale, PA, USA, 2013.
8. Nurliyana, C.; Lestari, Y.D.; Prasetyo, E.A.; Belgiawan, P.F. Exploring drivers' interest in different levels of autonomous vehicles: Insights from Java Island, Indonesia. *Transp. Res. Interdiscip. Perspect.* **2023**, *19*, 100820.
9. Islayem, R.; Alhosani, F.; Hashem, R.; Alzaabi, A.; Meribout, M. Hardware Accelerators for Autonomous Cars: A Review. *arXiv* **2024**, arXiv:2405.00062.
10. Hussain, M.; Hong, J.E. Reconstruction-Based Adversarial Attack Detection in Vision-Based Autonomous Driving Systems. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1589–1611. <https://doi.org/10.3390/make5040080>.
11. Azzaz, Z.; Khaled, O.; Khatab, E.; Said, H., and Shalash, O. EASE-PVNet: Robust Periocular Identity Verification Across Pre- and Post-Operative Facial Images. *Mach. Learn. Knowl. Extr.* **2026**, *8*, 169. <https://doi.org/10.3390/make8060169>.
12. Sana, F.; Azad, N.L.; Raahemifar, K. Autonomous vehicle decision-making and control in complex and unconventional scenarios—A review. *Machines* **2023**, *11*, 676.
13. Karras, A.; Theodorakopoulos, L.; Karras, C.; Theodoropoulou, A. Towards LLM-Driven Cybersecurity in Autonomous Vehicles: A Big Data-Empowered Framework with Emerging Technologies. *Mach. Learn. Knowl. Extr.* **2026**, *8*, 43. <https://doi.org/10.3390/make8020043>.
14. Khatab, E.; Onsy, A.; Varley, M.; Abouelfarag, A. Vulnerable objects detection for autonomous driving: A review. *Integration* **2021**, *78*, 36–48.
15. Shen, R.; Wang, Y.; Liu, H.; Gu, H.; Geng, C.; Shi, Y. Visual Perception and Robust Autonomous Following for Orchard Transportation Robots Based on DeepDIMP-ReID. *Mach. Learn. Knowl. Extr.* **2026**, *8*, 39. <https://doi.org/10.3390/make8020039>.
16. Pavone, M. How AI Is Unlocking Level 4 Autonomous Driving. NVIDIA Technical Blog, 2025. Available online: <https://blogs.nvidia.com/blog/level-4-autonomous-driving-ai/> (accessed on 23 January 2026).
17. Han, Y.; Zhang, H.; Li, H.; Jin, Y.; Lang, C.; Li, Y. Collaborative perception in autonomous driving: Methods, datasets, and challenges. *IEEE Intell. Transp. Syst. Mag.* **2023**, *15*, 131–151.
18. Chen, C.; Seff, A.; Kornhauser, A.; Xiao, J. Deepdriving: Learning affordance for direct perception in autonomous driving. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2722–2730.
19. Sun, C.; Zhang, R.; Lu, Y.; Cui, Y.; Deng, Z.; Cao, D.; Khajepour, A. Toward ensuring safety for autonomous driving perception: Standardization progress, research advances, and perspectives. *IEEE Trans. Intell. Transp. Syst.* **2023**, *25*, 3286–3304.
20. Marti, E.; De Miguel, M.A.; Garcia, F.; Perez, J. A review of sensor technologies for perception in automated driving. *IEEE Intell. Transp. Syst. Mag.* **2019**, *11*, 94–108.
21. Ignatious, H.A.; Khan, M. An overview of sensors in Autonomous Vehicles. *Procedia Comput. Sci.* **2022**, *198*, 736–741.
22. Yeong, D.J.; Velasco-Hernandez, G.; Barry, J.; Walsh, J. Sensor and Sensor Fusion Technology in Autonomous Vehicles: A Review. *Sensors* **2021**, *21*, 2140. <https://doi.org/10.3390/s21062140>.
23. Yurtsever, E.; Lambert, J.; Carballo, A.; Takeda, K. A survey of autonomous driving: Common practices and emerging technologies. *IEEE Access* **2020**, *8*, 58443–58469.
24. Lin, S.C.; Zhang, Y.; Hsu, C.H.; Skach, M.; Haque, M.E.; Tang, L.; Mars, J. The architectural implications of autonomous driving: Constraints and acceleration. In Proceedings of the Twenty-Third International Conference on Architectural Support for Programming Languages and Operating Systems, Williamsburg, VI, USA, 24–28 March 2018; pp. 751–766.
25. Liu, S.; Tang, J.; Zhang, Z.; Gaudiot, J.L. Computer architectures for autonomous driving. *Computer* **2017**, *50*, 18–25.
26. Shalash, O.; Emad, A.; Fathy, F.; Alzogby, A.; Sallam, M.; Naser, E.; El-Sayed, M.; Khatab, E. Fusion of Robotics, AI, and Thermal Imaging Technologies for Intelligent Precision Agriculture Systems. *Sensors* **2025**, *25*, 6844.
27. Sallam, M.; Salah, Y.; Osman, Y.; Hegazy, A.; Khatab, E.; Shalash, O. Intelligent Dental Handpiece: Real-Time Motion Analysis for Skill Development. *Sensors* **2025**, *25*, 6489.

28. Ayala, R.; Mohd, T.K. Sensors in autonomous vehicles: A survey. *J. Auton. Veh. Syst.* **2021**, *1*, 031003.
29. Vargas, J.; Alsweiss, S.; Toker, O.; Razdan, R.; Santos, J. An Overview of Autonomous Vehicles Sensors and Their Vulnerability to Weather Conditions. *Sensors* **2021**, *21*, 5397. <https://doi.org/10.3390/s21165397>.
30. Said, H.; Moniri, M.; Chibelushi, C.C. Prediction architecture based on block matching statistics for mixed spatial-resolution multi-view video coding. *EURASIP J. Image Video Process.* **2017**, *2017*, 15.
31. Shalash, O.; Sakr, A.; Salem, Y.; Abdelhadi, A.; Elsayed, H., and El-Shaer, A. Position and orientation analysis of Jupiter robot arm for navigation stability. *laes Int. J. Robot. Autom.* **2025**, *14*, 1–10, .
32. Said, H.; Akbari, A.S. H.264/AVC Based multi-view video codec using the statistics of block matching. In *Proceedings of the Proceedings ELMAR-2013*; IEEE: Piscataway, NJ, USA, 2013; pp. 97–100.
33. Tesla. Replacing Ultrasonic Sensors with Tesla Vision, 2025. Available online: https://www.tesla.com/en_jo/support/transitioning-tesla-vision (accessed on 23 February 2026).
34. Intel RealSense, Intel RealSense D455 Depth Camera. Available online: <https://realsenseai.com/stereo-depth-cameras/> (accessed on 26 June 2026).
35. Intel Corporation, Intel RealSense D435 Depth Camera. Available online: <https://store.intelrealsense.com/buy-intel-realsense-depth-camera-d435.html> (accessed on 26 June 2026).
36. Carnegie Robotics, MultiSense® KS21i Indoor Stereo Camera. Available online: <https://www.carnegierobotics.com/multisense-ks21i> (accessed on 26 June 2026).
37. Carnegie Robotics. MultiSense S27 Stereo Vision System. Available online: <https://shop.carnegierobotics.com/products/multisense-s27> (accessed on 26 June 2026).
38. Roboception GmbH. rc_visard 3D Stereo Sensor. Available online: <https://roboception.com/3d-stereo-vision/rc-visard-3d-stereo-sensor/> (accessed on 26 June 2026).
39. Stereolabs. ZED 2—AI Stereo Camera. Available online: <https://www.stereolabs.com/en-ae/products/zed-2> (accessed on 26 June 2026).
40. Stereolabs. ZED 2i Stereo Camera. Available online: <https://www.stereolabs.com/en-ae/store/products/zed-2i> (accessed on 26 June 2026).
41. Lu, X.; Wang, Y. *WeRide: Commercialization Exploration of an Autonomous Driving Technology Supplier*; Management School Fudan University: Shanghai, China, 2024; pp. 1–23.
42. Hennaoui, H.; Paluszczyszyn, D.; Deka, L.; Cosar, S. A Framework for Assessment of Perception Systems in Autonomous Vehicles. *IEEE Access* **2025**, *13*, 110900–110912.
43. Sharif, D.; Murtala, S.; Choi, G.S. A Survey of Automotive Radar Misalignment Detection Techniques. *IEEE Access* **2025**, *13*, 123314–123324. <https://doi.org/10.1109/ACCESS.2025.3584454>.
44. Jusoh, S.; Almajali, S. Sensor Fusion Technology Advancement in GPS-Aided Localization for Autonomous Mobile Robots: A Comprehensive Survey. *Jurnal Teknologi* **2025**, *87*, 165–189.
45. Rajashekara, K.; Koppera, S. Data and energy impacts of intelligent transportation—A review. *World Electr. Veh. J.* **2024**, *15*, 262.
46. Khattab, Y.; Zidane, I.F.; El-Habrouk, M.; Rezek, S. Solving kinematics of a parallel manipulator using artificial neural networks. In *Proceedings of the 2021 31st International Conference on Computer Theory and Applications (ICCTA)*; IEEE: Piscataway, NJ, USA, 2021; pp. 84–89.
47. Fayyad, J.; Jaradat, M.A.; Gruyer, D.; Najjaran, H. Deep Learning Sensor Fusion for Autonomous Vehicle Perception and Localization: A Review. *Sensors* **2020**, *20*, 4220. <https://doi.org/10.3390/s20154220>.
48. Matos, F.; Bernardino, J.; Durães, J.; Cunha, J. A survey on sensor failures in autonomous vehicles: Challenges and solutions. *Sensors* **2024**, *24*, 5108.
49. Khatab, E.; Onsy, A.; Varley, M.; Abouelfarag, A. A lightweight network for real-time rain streaks and rain accumulation removal from single images captured by avs. *Appl. Sci.* **2022**, *13*, 219.
50. Khatab, E.; Onsy, A.; Abouelfarag, A. Evaluation of 3D vulnerable objects’ detection using a multi-sensors system for autonomous vehicles. *Sensors* **2022**, *22*, 1663.
51. Prakash, A.; Chitta, K.; Geiger, A. Multi-modal fusion transformer for end-to-end autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 19–25 June 2021; pp. 7077–7087.
52. Abouelfarag, A.; El-Shenawy, M.; Khatab, E. High speed edge detection implementation using compressor cells over rsda. In *Proceedings of the International Conference on Interfaces and Human Computer Interaction 2016, Game and Entertainment Technologies 2016 and Computer Graphics, Visualization, Computer Vision and Image Processing 2016-Part of the Multi Conference on Computer Science and Information Systems 2016*; IADIS Press: Lisbon, Portugal, 2016; pp. 206–214.
53. Feng, X.; Jiang, Y.; Yang, X.; Du, M.; Li, X. Computer vision algorithms and hardware implementations: A survey. *Integration* **2019**, *69*, 309–320.
54. Asvadi, A.; Girao, P.; Peixoto, P.; Nunes, U. 3D object tracking using RGB and LIDAR data. In *Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*; IEEE: Piscataway, NJ, USA, 2016; pp. 1255–1260.

55. Akbari, A.S.; Said, H.; Moniri, M. Effect of inter-camera angles on the performance of an H. 264/AVC based multi-view video codec. In *Proceedings of the 2012 Picture Coding Symposium*; IEEE: Piscataway, NJ, USA, 2012; pp. 109–112.
56. Cui, H.; Radosavljevic, V.; Chou, F.C.; Lin, T.H.; Nguyen, T.; Huang, T.K.; Schneider, J.; Djuric, N. Multimodal trajectory predictions for autonomous driving using deep convolutional networks. In *Proceedings of the 2019 International Conference on Robotics and Automation (ICRA)*; IEEE: Piscataway, NJ, USA, 2019; pp. 2090–2096.
57. Geisberger, R.; Sanders, P.; Schultes, D.; Vetter, C. Exact routing in large road networks using contraction hierarchies. *Transp. Sci.* **2012**, *46*, 388–404.
58. Sanders, P.; Schultes, D. Highway hierarchies hasten exact shortest path queries. In *Proceedings of the European Symposium on Algorithms*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 568–579.
59. Goldberg, A.V.; Kaplan, H.; Werneck, R.F. Reach for A*: Efficient point-to-point shortest path algorithms. In *Proceedings of the Eighth Workshop on Algorithm Engineering and Experiments (ALENEX)*; SIAM: Philadelphia, PA, USA, 2006; pp. 129–143.
60. Pivtoraiko, M.; Kelly, A. Efficient constrained path planning via search in state lattices. In *Proceedings of the 8th International Symposium on Artificial Intelligence, Robotics and Automation in Space*, Munich, Germany, 5–8 September 2005.
61. Chitta, K.; Prakash, A.; Jaeger, B.; Yu, Z.; Renz, K.; Geiger, A. Transfuser: Imitation with transformer-based sensor fusion for autonomous driving. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 12878–12895.
62. Sahoo, L.K.; Varadarajan, V. Deep learning for autonomous driving systems: Technological innovations, strategic implementations, and business implications—a comprehensive review. *Complex Eng. Syst.* **2025**, *5*, 12, doi:10.20517/ces.2024.83.
63. Zhang, T.; Liu, H.; Wang, W.; Wang, X. Virtual tools for testing autonomous driving: A survey and benchmark of simulators, datasets, and competitions. *Electronics* **2024**, *13*, 3486.
64. Shah, S.; Dey, D.; Lovett, C.; Kapoor, A. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Proceedings of the Field and Service Robotics: Results of the 11th International Conference*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 621–635.
65. Dosovitskiy, A.; Ros, G.; Codevilla, F.; Lopez, A.; Koltun, V. CARLA: An open urban driving simulator. In *Proceedings of the Conference on Robot Learning*, Mountain View, CA, USA, 13–15 November 2017; pp. 1–16.
66. Baidu Apollo. Apollo: Open Autonomous Driving Platform. Apollo Developer Community Website, 2026. Available online: <https://apollo.baidu.com/> (accessed on 7 January 2026).
67. The Autoware Foundation. Autoware. The Autoware Foundation Website, 2026. Available online: <https://autoware.org/> (accessed on 7 January 2026).
68. Koenig, N.; Howard, A. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE Cat. No. 04CH37566; IEEE: Piscataway, NJ, USA, 2004; Volume 3, pp. 2149–2154.
69. Li, W.; Pan, C.; Zhang, R.; Ren, J.; Ma, Y.; Fang, J.; Yan, F.; Geng, Q.; Huang, X.; Gong, H.; et al. AADS: Augmented autonomous driving simulation using data-driven algorithms. *Sci. Robot.* **2019**, *4*, eaaw0863.
70. Yao, S.; Zhang, J.; Hu, Z.; Wang, Y.; Zhou, X. Autonomous-driving vehicle test technology based on virtual reality. *J. Eng.* **2018**, *2018*, 1768–1771.
71. Sovani, S. Simulation accelerates development of autonomous driving. *ATZ Worldw.* **2017**, *119*, 24–29.
72. Cognata. Autonomous and ADAS Vehicles Simulation. Cognata Official Website, 2026. Available online: <https://www.cognata.com/autonomous-vehicles/> (accessed on 7 January 2026).
73. Deemantha, R.; Hettige, B. Autonomous car: Current issues, challenges and solution: A review. In *Proceedings of the IEEE International Conference on Intelligent Transportation Systems, ITSC 2023*, Bilbao, Spain, 24–28 September 2023.
74. Gajjar, H.; Sanyal, S.; Shah, M. A comprehensive study on lane detecting autonomous car using computer vision. *Expert Syst. Appl.* **2023**, *233*, 120929.
75. Fahadullah, F.; Saeed, R. WAYMO AND V2V: BRIDGING THE GAP BETWEEN AUTONOMOUS AND HUMAN-DRIVEN VEHICLES. *J. Mind Behav.* **2025**. Available online: https://www.researchgate.net/profile/Fnu-Fahadullah/publication/396886557_WAYMO_AND_V2V_BRIDGING_THE_GAP_BETWEEN_AUTONOMOUS_AND_HUMAN-DRIVEN_VEHICLES/links/694ca3939aa6b4649dc46aaa/WAYMO-AND-V2V-BRIDGING-THE-GAP-BETWEEN-AUTONOMOUS-AND-HUMAN-DRIVEN-VEHICLES.pdf (accessed on 23 January 2026).
76. Waymo LLC. Waymo Safety Report. *Company Safety Report*, 2021. Available online: <https://waymo.com/research/waymo-safety-report/> (accessed on 23 January 2026).
77. Sjöberg, K. Robotaxis Will Always Need People [Connected and Automated Vehicles]. *IEEE Veh. Technol. Mag.* **2025**, *20*, 135–137.
78. Wang, S.; Zhao, Z.; Xie, Y.; Ma, M.; Chen, Z.; Wang, Z.; Su, B.; Xu, W.; Li, T. Recent surge in public interest in transportation: Sentiment analysis of Baidu Apollo Go using Weibo data. *arXiv* **2024**, arXiv:2408.10088.
79. Min, J.; Hong, Y.; King, C.B.; Meeker, W.Q. Reliability analysis of artificial intelligence systems using recurrent events data from autonomous vehicles. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2022**, *71*, 987–1013.
80. Pony.ai. Technology, 2025. Available online: <https://ir.pony.ai/> (accessed on 7 December 2025).

81. Garikapati, D.; Shetiya, S.S. Autonomous vehicles: Evolution of artificial intelligence and the current industry landscape. *Big Data Cogn. Comput.* **2024**, *8*, 42.
82. Waymo, L. Introducing the 5th-Generation Waymo Driver: Informed by Experience, Designed for Scale, Engineered to Tackle More Environments, 2020. Available online: <https://waymo.com/blog/2020/03/introducing-5th-generation-waymo-driver/> (accessed on 4 March 2020).
83. Ram, G.S.S. Waymo's AI and Robotic Architecture: A Deep Dive with Novel Prediction Enhancements. Available online: <https://www.techrxiv.org/doi/full/10.36227/techrxiv.174742706.68651162> (accessed on 16 May 2025).
84. Chen, L.; Wu, P.; Chitta, K.; Jaeger, B.; Geiger, A.; Li, H. End-to-end autonomous driving: Challenges and frontiers. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 10164–10183.
85. Bojarski, M.; Del Testa, D.; Dworakowski, D.; Firner, B.; Flepp, B.; Goyal, P.; Jackel, L.D.; Monfort, M.; Muller, U.; Zhang, J.; et al. End to end learning for self-driving cars. *arXiv* **2016**, arXiv:1604.07316.
86. Salah, Y.; Shalash, O.; Khatab, E. A lightweight speaker verification approach for autonomous vehicles. *Robot. Integr. Manuf. Control* **2024**, *1*, 15–30.
87. ISO 26262-2:2018; Road Vehicles—Functional Safety—Part 2: Management of Functional Safety. ISO: Geneva, Switzerland, 2018.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.