



# FISITA

Promoting excellence in mobility engineering

iStock™

Credit: Just\_Super

## Automated Driving Expert Group White Paper

The Data Imperative: Safety assurance for automated driving through data collection and analysis



[www.fisita.com](http://www.fisita.com)



[www.linkedin.com/company/FISITA](http://www.linkedin.com/company/FISITA)

# Contents

|    |                                                                             |
|----|-----------------------------------------------------------------------------|
| 2  | <a href="#">Executive Summary</a>                                           |
| 3  | <a href="#">Introduction</a>                                                |
| 3  | <a href="#">A brief history of automated driving</a>                        |
| 3  | <a href="#">Automated highway systems</a>                                   |
| 4  | <a href="#">Vehicle-based automation</a>                                    |
| 6  | <a href="#">Current regulatory landscape for automated driving safety</a>   |
| 7  | <a href="#">The importance of public trust</a>                              |
| 8  | <a href="#">Use of human driving as a benchmark for automated vehicles</a>  |
| 10 | <a href="#">The role of data in demonstrating safety and building trust</a> |
| 11 | <a href="#">Challenge of end-to-end deep learning</a>                       |
| 13 | <a href="#">Conclusion</a>                                                  |
| 14 | <a href="#">Contributors</a>                                                |
| 15 | <a href="#">References</a>                                                  |

## Executive Summary

Automated driving has evolved over nearly a century, from early infrastructure-guided prototypes to today's vehicle-centric, AI-enabled systems. Each stage in this development has been driven by the pursuit of safe, clean, efficient and accessible mobility yet has also raised fundamental questions about how the safety of automated vehicles (AVs) can be assured, demonstrated and trusted.

This paper, produced by the FISITA Automated Driving Expert Group, traces the historical progression of automated driving, outlines the current regulatory and standards landscape and explores how data can underpin credible, evidence-based safety assurance. As automated driving technologies transition from rule-based architectures to complex, end-to-end deep learning systems, conventional approaches to proving safety through accumulated mileage or human-driver comparisons are no longer sufficient. Instead, the assurance process must shift toward continuous, data-driven evaluation.

AVs generate multiple categories of data that are essential for this process:

- Operational data captures vehicle dynamics, system status and environmental conditions.
- Perception and decision-making data records how the vehicle senses, interprets and responds to its surroundings.
- Event data provides detailed context for incidents, disengagements and near misses.
- Training data supports the development and refinement of machine learning algorithms.

Together, these datasets offer a rich source of insight into system performance but their regulatory use is complicated by the diversity of hardware, software and data architectures among developers. The absence of standardised formats and common interpretive frameworks limits transparency, comparability and effective oversight.

To address these challenges, the British Standards Institution's (BSI) Digital Commentary Driving (DCD) framework proposes a standardised method for capturing and communicating how AVs perceive and decide, inspired by the human practice of commentary driving. By defining a consistent structure for explaining system reasoning, DCD provides data that enables examination of AV perceptions, predictions and behaviours to assess safety performance without exposing proprietary code, thereby enabling more transparent regulatory evaluation and public accountability.

This paper concludes that data transparency is both a technical and societal imperative. Building public trust in automated driving will depend on establishing shared frameworks for data collection, interpretation and communication, ensuring that automated vehicles not only operate safely but can be seen to do so. At the same time, consumers may need to accept confidential collection and sharing of data of their use of AVs to support continuous improvement of the safety and performance of such services. Through responsible data use and collaborative safety assurance, the vision of safe, clean, efficient and accessible mobility can move from aspiration to reality.

# Introduction

## A brief history of automated driving

### Automated highway systems

Automated driving has been promised for decades. After early initial experiments in radio control of vehicles (Houdina's 'American Wonder' of the 1920s; Engelking, 2017), a compelling vision for automated driving was captured by GM's 'Futurama' exhibit at the 1939 World Fair in New York. Conceived by famed industrial designer Norman Bel Geddes, this presented a conceptual transport system with U.S. cities connected by multi-lane highways upon which vehicles were automatically guided at high speed while vehicle occupants were free to engage in other activities (Bel Geddes, 1940). The ability to complete high-speed, direct, safe, intercity journeys whilst the driver is freed from the requirement to attend to the driving task is a concept that inspired the development of automated driving technologies over the remainder of the century.

Attempts to realise Bel Geddes' vision using guidance equipment integrated into physical infrastructure, such as magnetic beacons or buried electric cables, were successfully tested (see Figure 1). However, the projected installation and maintenance costs for such systems to achieve mass adoption were significant with considerable non-technical issues (e.g. liability, socioeconomic impacts, stakeholder alignment etc.) to address (Cheon, 2003). Ultimately, aligning the challenges of technological feasibility with the economic realities of an infrastructure-based approach to automated driving proved to be insurmountable. As computing and sensing systems grew in capability over the second half of the 20th century, automated driving research therefore shifted to embedding vehicles with the technology needed for safe, effective guidance.



*Figure 1. Citroen DS fitted with sensors and systems that enabled automated driving by following an electric guidance cable buried in the test track at the UK's Road Research Laboratory (picture credit: TRL Limited).*

## Vehicle-based automation

Ernst Dickmanns and colleagues working at Universität der Bundeswehr München (University of the Bundeswehr Munich) developed early computer vision techniques using Kalman filters that would create the foundations for automated driving based around in-vehicle technology for monitoring road environments and adapting vehicle behaviour in response (see Dickmanns & Zapp, 1987). This led to significant funding from the European Union in the form of a collaborative R&D project called PROMETHEUS (Programme for a European Traffic of Highest Efficiency and Unprecedented Safety, 1987-1995). Led by Daimler-Benz and with partners including BMW, Peugeot, Renault, Fiat, Volvo, Bosch, Siemens, and research institutions and universities from across Europe, the PROMETHEUS project developed early forms of computer vision, sensor fusion and dynamic trajectory planning and achieved the first high-speed, fully automated driving on public highways (Dickmanns, 2002).

Contemporary work (late 1980s to mid 1990s) was being carried out at Carnegie Mellon University's Robotics Institute in their Navlab program. Their work was supported by funding from traditional academic and military sources, and this guided the direction of their research. Starting with a Chevrolet van (Navlab 1) equipped with cameras, LIDAR, radar and ultrasonic sensors and onboard computers, they progressed to a similar set-up using a modified U.S. Army High Mobility Multipurpose Wheeled Vehicle (HMMWV or 'Humvee'; Navlab 2). This latter vehicle was able to operate at higher speeds and in more complex, rugged environments to suit potential military applications. Further Navlab vehicles were developed, eventually shifting to civilian cars and buses and were used to test an artificial neural network approach to automated driving (ALVINN – Autonomous Land Vehicle In a Neural Network; see Pomerleau, 1989). Simulated camera and laser inputs from a vehicle were used as the input to train a three-layer back-propagation artificial neural network and the output from the network was the direction of travel that the vehicle should follow. This was shown to work successfully in the Navlab vehicles under certain conditions, but it was recognised that greater variety was needed in the training dataset to give the system greater robustness for real world driving (ibid.).

The achievements of PROMETHEUS and Navlab gave encouragement to the idea that vehicle-based automated driving was achievable, and military applications were seen as a potential priority. One tragic feature of conflicts around the turn of the century was the effectiveness of improvised explosive devices (IEDs) in disrupting or destroying military vehicles, causing fatalities and serious injury to service personnel and civilians. Indeed, road crashes in general can be a significant cause of the loss of military personnel with (typically) high risk, young, male drivers responsible for vehicle movements between bases (Parker, 2023). With such risks in mind, it was recognised that automated driving could enable the safe movement of cargo-carrying vehicles and reduce the threat to human life posed by IEDs. The U.S. defence research agency, DARPA (Defense Advanced Research Projects Agency) runs 'Challenge' events in which research consortia are invited to compete in developing and demonstrating technologies that could enhance defence capabilities. To explore the potential for automation of military land vehicles and reduce the potential for harm to personnel, DARPA held a series of Challenges in the 2000s with the objective to enable a vehicle to complete a designated journey without input from a driver or operator. With routes gradually increasing in complexity, the entries for the 2004, 2005, and 2007 Challenges essentially set the template for how to build a vehicle capable of self-driving (see Figure 2). They also created a cohort of researchers that would go on to establish many of the leading organisations developing automated vehicle technology. Notable examples include Sebastian Thrun, Chris Urmson, and Bryan Salesky and DARPA Challenge alumni taking leadership roles in the Google Self-Driving Car Project / Waymo, Cruise, Aurora, Argo AI, Uber ATG, and Nuro.



*Figure 2. Winner of the 2007 DARPA Urban Challenge (2007) – a collaborative effort by Carnegie Mellon University and General Motors Corporation, a heavily modified Chevrolet Tahoe, nicknamed 'Boss' (picture credit: N. Reed).*

Compared to vehicles intended to operate on automated highways, the DARPA Challenge vehicles featured a dramatic increase in the number and variety (cameras, lidar, radar) of sensors fitted to the vehicle in order to collate the necessary and sufficient information about its surroundings and to enable computation of the appropriate driving behaviours. With multiple sensor types providing coverage around the vehicle, the limitations of one could be overcome by another (e.g. radar sensors detecting obstacles when cameras have been obscured by fog). However, while automated driving could be achieved in the real world within certain constraints using this approach, it typically depended on the use of hard coded rules of operation and up-to-date, detailed digital maps of the operating environment. This specificity made the task of scaling up automated operations more challenging as each new environment or feature of the intended operational design domain (ODD) potentially required significant updates to the automated driving system.

By contrast, experienced human drivers are capable of driving effectively in a wide variety of environments and making intuitive decisions about how to resolve unfamiliar scenarios safely using a rather limited sensor set (predominantly our vision but supported by useful information from our auditory, haptic and proprioceptive senses) and cognitive computation. Inspired by this approach, other developers – particularly Tesla and Wayve – have been collecting driving data and training artificial neural networks in how to drive using data collected by a simpler and cheaper sensor payload than that used by others.

Rather than developing a traditional modular approach to the tasks of perception and planning, a complex 'deep learning' system receives sensor input and is trained to output driving behaviours – a so-called end-to-end deep learning approach. Critically, the intention of this approach is to create a system that has derived the principles of safe driving from its training data and can then apply those in essentially any environment, thereby making it far quicker, cheaper and easier to scale automated driving capabilities to multiple locations.

However, an important limitation in the use of end-to-end deep learning systems is the challenge of explainability – understanding why the vehicle produces the behaviours that are observed. This makes it more

difficult to prove to a regulator that the technology will always be able to drive in an acceptably safe manner when confronted by the infinite complexity that it may encounter in real world driving. The section that follows discusses the current regulatory landscape for automated vehicles and the areas where further work is required to build regulatory frameworks that enable safe, trusted operation of automated vehicles.

In this paper, we consider the regulations that apply to automated driving at present and explore why achieving public trust, as a fundamental pillar of commercial adoption, will demand a collaborative approach to data sharing.

## Current regulatory landscape for automated driving safety

The global regulatory landscape for automated vehicles (AVs) is converging towards similar safety principles but shaped by distinct regional philosophies. In the United States, regulation is predominantly market-driven and decentralised with innovation led by industry and oversight split between federal and state-level guidance. Bodies such as the National Highway Traffic Safety Administration and the Federal Highway Administration issue non-binding guidance and exemptions rather than prescriptive approval regimes, allowing rapid deployment but resulting in fragmented requirements across states. China, by contrast, has adopted a highly centralised national strategy. Central government sets strategic direction while city and provincial authorities authorise large-scale pilots, enabling rapid learning and deployment but within a tightly controlled framework. Europe and the UK are pursuing a comprehensive type-approval-based approach in which automated driving functions must demonstrate compliance against harmonised international regulations and standards before market entry.

At the heart of this harmonised approach is the work of the United Nations Economic Commission for Europe (UNECE), whose World Forum for Harmonization of Vehicle Regulations (WP.29) has become the global focal point for AV regulation. Through WP.29 and its GRVA working group, UNECE has established binding UN Regulations that many jurisdictions adopt directly into national law. Notably, UN Regulation No. 155 mandates a lifecycle Cybersecurity Management System, and UN Regulation No. 156 requires a Software Update Management System to ensure that software changes are safe, auditable, and controlled throughout a vehicle's life. UN Regulation No. 157 on Automated Lane Keeping Systems (ALKS) marked the first internationally agreed framework for SAE Level 3 automation, introducing requirements for driver monitoring, operational design domain constraints, and mandatory data recording via a Data Storage System for Automated Driving (DSSAD). Together, these regulations signal a shift from one-off vehicle approval toward continuous, data-supported safety assurance. Building on this, UNECE launched a major initiative in 2024 to develop a comprehensive global framework for Automated Driving Systems, with both a UN Regulation and a Global Technical Regulation on ADS expected in 2026 (adopted in January 2026).

These regulations are underpinned by an expanding ecosystem of international standards, particularly from ISO and SAE. ISO 26262 (functional safety) and ISO/SAE 21448 (Safety of the Intended Functionality, SOTIF) address hazards arising from system failures and performance limitations, while newer standards such as ISO/SAE PAS 21450 (ADS safety case), ISO 22737 (low-speed automated driving), and ISO 34502 (scenario-based testing) increasingly formalise how safety arguments are structured, evidenced, and validated.

EU Regulation 2022/1426 establishes uniform procedures and technical specifications for the type-approval of Automated Driving Systems (ADS) in fully automated vehicles under EU law, supplementing the general safety rules of Regulation (EU) 2019/2144. It defines specific use-cases (predefined areas, fixed routes, automated parking) and the technical and assessment requirements manufacturers must meet for compliance. This

framework aims to harmonize safety standards and enable the introduction of fully automated vehicles to the European market.

National authorities layer additional guidance on top of this foundation, for example Germany's BASt guidance on automated driving approval, Japan's MLIT frameworks for Level 3 deployment, and evolving Chinese national and municipal rules, while the European Union adds horizontal regulation through instruments such as the General Safety Regulation and the AI Act, extending assurance requirements beyond vehicles to algorithms, data governance, and lifecycle accountability.

Across all regions, a common challenge is emerging – how to move from prescriptive, function-specific rules to scalable, data-driven safety assurance capable of addressing complex real-world behaviour and the “long tail” of rare but critical scenarios. Regulators and industry alike face issues of data volume, representativeness, interoperability, acceptable risk definition, and secure sharing. At the same time, there is a clear opportunity to leverage large-scale operational data, simulation, and AI-enabled analysis to support continuous learning and regulatory confidence. The direction of travel is toward globally aligned frameworks that combine international regulation, recognised standards, and transparent, evidence-based safety cases. These should enable innovation while maintaining public trust and regulatory oversight as automated driving systems move from trials to pilots to widespread commercial deployment.

## The importance of public trust

Public trust is crucial for the widespread acceptance and successful integration of automated driving systems, providing a ‘social licence’ to operate. People need to feel confident that automated vehicles are at least as safe, if not safer, than human-driven vehicles and this factor is frequently cited as a critical concern in public surveys of automated vehicle adoption (e.g. Kyriakidis, Happee & de Winter, 2015; Kaur & Rampersad, 2018). Examples such as genetically modified foods (Frewer, Scholderer, & Bredahl, 2003) and attitudes to nuclear power after the Fukushima disaster (Kim, Kim & Kim, 2013) show that if trust in a technology is compromised, it may meet significant resistance from the public. In essence, public trust is the foundation upon which the future of automated driving rests. Without it, the potential benefits of this technology may never be fully realised.

The need for automated vehicle deployments to improve safety was captured in the European Commission's recommendations on the ethics of connected and automated vehicles (Bonnefon et al., 2020) and such principles are now being encapsulated in legislation governing the operation of automated vehicles (e.g. the UK Automated Vehicles Act 2024, United Kingdom, 2024). Any incidents involving automated vehicles can severely erode public trust. Both Uber and Cruise ultimately closed their automated vehicle development programmes following high-profile incidents (see Korosec, 2020, 2023, 2024). People need to trust that the technology will perform reliably in various real-world scenarios, including in challenging and unexpected situations.

Automated vehicle developer, Waymo, has probably published the most material of any company involved in the sector to demonstrate the safety of its vehicles (e.g. Favaro et al., 2025; Fraade-Blanar et al., 2025; Kusano et al., 2025; see <https://waymo.com/safety/>). Waymo's position as a leader in establishing the safety of its vehicles has its foundation in the millions of miles of automated operation on public roads (Kusano et al., 2025). However, even this incredible achievement fails to reach the statistical benchmark envisaged by a research paper from the RAND Corporation (Kalra & Paddock, 2016) which estimated that:

*“Autonomous vehicles would have to be driven hundreds of millions of miles and sometimes hundreds of billions of miles to demonstrate their reliability in terms of fatalities and injuries.”*  
(Kalra & Paddock, 2016)

Although there are more road crashes than is acceptable, they are still relatively rare in relation to the amount of driving that takes place. This is why Kalra and Paddock’s estimate of the amount of driving needed by automated vehicles to be able to achieve statistical confidence that they are safer than human drivers is so high.

The challenge for the industry therefore is to develop tools and techniques that could enable automated vehicle developers to show that their technology is operating safely in every drive and thereby give assurance to both regulators and the public that deployments are acceptable without needing hundreds of millions of miles of data to achieve statistical proof of safety performance.

## Use of human driving as a benchmark for automated vehicles

One approach to finding an acceptable level of performance for automated driving safety is to prove that automated vehicles can drive with a level of safety that exceeds that of human drivers. With established casualty statistics for human driving and existing traffic laws and infrastructure based around human drivers, a reduction in the rate of serious incidents for automated vehicles could be seen as an acceptable level of safety performance.

However, human drivers are prone to risks from distraction, fatigue and impairment, which contribute significantly to crash causation. While automated vehicles will not experience the same risks, they may have different failure modes. These may be seen as less acceptable by the public. Recent studies (e.g. Waymo and Swiss Re – Di Lillo et al., 2024) are using insurance claims data to provide a more robust comparison between AV and human driving safety. The Waymo / Swiss Re study suggests that AVs can perform significantly better than human drivers with lower claims frequency and incident severity. However, whilst the authors apply necessary diligence over the statistical power of their analyses, care is needed to ensure sufficient data is collated in order to achieve fair comparisons and to ensure that the human data being used for comparison represents a suitable benchmark. For example, if an AV fleet is only capable of highway driving, its safety should not be compared to human driving in urban or rural environments, which typically have significantly higher collision rates.

It is also important to consider failure modes. The types of collision involving AVs might be very different to those involving human vehicles. If an AV fleet normally displays exemplary driving behaviour but occasionally causes preventable crashes due to unreliable sensors, it is unlikely to achieve approval to operate without a safety driver.

Finally, a valid approach to assessing AV safety would be to compare how AVs perform in simulations of real world driving and compare their performance to that observed of human drivers. This is included for example, in the UN Regulation 157 which includes a demonstration of the driver model used for the ALKS safety concept. Simulation is an excellent way to control many of the variables that affect our ability to evaluate driving performance under real world conditions and to drill down into AV driving behaviours in the virtual environment. Further, simulation offers the potential to increase the value of each mile driven in evaluating AV safety performance. Driving an AV on low complexity, low traffic roads in good weather would

enable a developer to rack up significant distance but without adding huge value in terms of validating safety performance. In a simulated world, a wide variety of hazards and conditions can be presented in succession enabling more efficient exploration of the problem space.

In a synthetic world, the 'ground truth' of the locations, movements and behaviours of all traffic participants can be measured objectively to assess safety performance. However, to be able to use simulations as part of a safety case for AV deployment in the real world, it must be possible to quantify the fidelity of the models used for simulation purposes. This has become an optimisation challenge, where the level of granularity and representation can change according to the system under test priorities (Leahy et al., 2026).

Several factors such as component wear, precipitation / spray effects, lens flare, reflections, particulates, and repeated power cycles may influence an AV's driving behaviour in the real world. Similarly, the behaviours and appearance of other road users, including relevant pedestrian factors (e.g. age range, gender, disability, clothing types, etc.), vehicle types (e.g. bicycle, skateboard, horse, e-scooter, moped, tuk-tuk, large truck, etc.) and actions (e.g. pedestrian stepping into the road without looking, cyclist indicating using hand signals, motorcycle courier filtering through traffic, etc.) should all be represented at a level of fidelity that ensures that any assessment of AV safety reflects performance in the real world.

A further challenge for simulation is to model interactions between the automated entity and other agents. For example, how do pedestrians respond to the behaviour of the AV, and can the AV then adapt its behaviour in response? Can interactions between AVs and emergency responders be managed within simulations and truly reflect the often safety- and time-critical nature of such interactions? Such questions highlight the need for multiple strands of safety assurance in addition to simulation data.

## The role of data in demonstrating safety and building trust

Data plays a pivotal role in building public trust in the safety of automated driving systems by providing tangible evidence and transparency over the safety of automated vehicle operations. Data allows for the quantification of safety performance. Metrics like disengagement rates, collision rates and near-miss incidents can be tracked and analysed. This provides evidence of how well the systems are performing, in support of or to contradict subjective opinions.

Data from simulations and real-world testing enables rigorous validation and verification of safety prior to full deployment. This demonstrates that the systems have been thoroughly tested under various conditions. Data from monitoring of AVs in-use and particularly following incidents can also be used to detect safety concerns and identify areas for improvement. Sharing this data with regulators and the public can increase transparency and demonstrate a commitment to safety. In surveys and workshops focused on AV ethics, Reed et al. (2023) reported that there is a public expectation of data sharing with approved, authorised stakeholders in the interests of improving safety.

AV data streams fall broadly into three categories: operational data, perception and decision-making data, and event data:

- **Operational data** includes information about vehicle dynamics (speed, acceleration, braking, steering inputs), system status (sensor health, software version, fault codes) and environmental conditions (weather, road surface, lighting). These datasets allow developers and regulators to understand how the vehicle behaves under varying real-world conditions and whether it remains within safe operational limits.
- **Perception and decision-making data** arises from the vehicle's sensor suite and control stack, including camera, radar, lidar, ultrasonic and GPS inputs, as well as the internal representations generated by perception, prediction, and planning algorithms. These enable a detailed reconstruction of how the AV 'saw' and interpreted the world around it, what options it considered and why it chose a particular course of action
- **Event data**, similar in principle to an aviation flight recorder, captures information leading up to, during, and after significant occurrences such as disengagements, near-misses, or collisions. This data is invaluable for root-cause analysis, enabling developers and investigators to determine whether system faults, external factors, or unexpected behaviours led to safety-critical outcomes.

Together, these datasets can be used to evaluate safety performance, validate system updates, inform simulation models and provide evidence to regulators that the technology behaves safely within its intended operational design domain.

However, using such data for regulatory purposes presents major challenges. The diversity of AV developers, from established automotive manufacturers to start-ups, means that sensing configurations, data architectures and software pipelines differ widely. Some companies use heavy sensor redundancy with detailed 3D mapping and rule-based decision trees, while others rely on camera-only, end-to-end deep learning systems trained on millions of human-driving examples. As a result, data formats, recording periods, sampling rates, metadata, and annotation methods vary significantly, making it difficult for regulators to compare performance objectively across developers or to define consistent safety metrics. Proprietary data handling and commercial confidentiality compound the issue, often restricting access to the very evidence required for safety validation. Furthermore, terms such as 'control', 'disengagement', 'intervention', and 'near-miss' have

different meanings to different stakeholders, raising the risk of misunderstandings and misleading conclusions about system behaviour and responsibility in safety-critical events.

The British Standards Institution (BSI) proposed the Digital Commentary Driving (DCD) approach, which offers a promising way to overcome some of these challenges. Inspired by the human practice of commentary driving, where drivers verbally describe hazards, intentions, and reasoning while driving, DCD provides a structured, standardised method for capturing and annotating an AV's perception, prediction, and decision processes. Rather than prescribing a particular hardware or software architecture, DCD defines a common framework for how systems record, describe, and communicate their situational understanding.

By standardising how driving behaviour and reasoning are expressed, DCD enables consistent interpretation of safety-relevant data across different AV platforms. It allows developers to demonstrate why the vehicle behaved as it did in any given situation, supports independent audit and comparison, and facilitates regulator understanding without requiring full access to proprietary code or datasets. In this way, DCD bridges the gap between the diversity of technical approaches and the regulatory need for transparent, comparable, and evidence-based safety assurance.

## Challenge of end-to-end deep learning

Developing AVs based solely on hard-coded rules of operation can work for specific, defined use cases but this approach has a brittleness that limits opportunities for commercialisation. AVs that use an end-to-end deep learning approach overcome this challenge as they are more capable of applying what they have determined to be the underlying principles of safe driving in different environments and applications. In 2025, the AV developer, Wayve, showcased the capabilities of its AI-based approach by demonstrating that its vehicle could drive in 500 cities without any retraining (Wayve, 2025).

However, building transparency into processes delivered by end-to-end deep learning systems is a significant and ongoing challenge. These systems, by their nature, often operate as 'black boxes'. Deep learning models, especially large neural networks, can have millions or even billions of parameters learning complex, non-linear relationships within the data, which are often not easily interpretable by humans. This complexity makes it extremely difficult to trace the exact reasoning behind a particular output and when the model is updated to fix a particular issue, it can be hard to know whether or not a new issue has arisen in behaviours that were previously deemed acceptable. In 2025, Bosch announced a modular end-to-end deep learning approach for its latest ADAS technology (see Bosch, 2025). This uses deep learning techniques but within defined modules in the ADAS system. It is claimed that this balances the controllability and interpretability of traditional approaches with flexibility and performance of a full end-to-end approach (Bosch, 2025).

Aside from traditional crashworthiness metrics for its vehicles, Tesla's public approach to proving the safety of its Autopilot driver assistance technology uses crash rate comparisons, with Tesla vehicles using Autopilot compared against Tesla vehicles not using Autopilot and against the U.S. national average (see Tesla, 2025). Wayve has not deployed AVs without safety drivers to date but has been working on how it will achieve safety assurance. Its DriveSafeAI project, in partnership with Warwick Manufacturing Group (WMG), suggests a focus on operational design domain safety assurance methods, simulation tools and AV architecture validation. While Tesla and Wayve might achieve statistical assurance over the safety performance of their technology, their approaches seem to leave a gap in providing detail around how and why incidents involving their vehicles occur and the associated risk of losing public trust.

The International AI Safety Report 2026 (Bengio et al., 2026) observes that AI capabilities are advancing rapidly, particularly due to post-training optimisation methods. Systems now perform at expert level in complex reasoning tasks across mathematics, software engineering, and science, and can generate high-quality code and photorealistic images. Yet they still struggle with basic abilities such as reliably counting objects, reasoning about physical space, or recovering from simple workflow errors. Evidence of AI-related risks is also increasing, notably in the areas of malicious use, accidental malfunctions, and systemic vulnerabilities; all of these are issues with significant implications for safety-critical domains. The central question is how these risks can be effectively mitigated.

A layered risk-management approach is widely advocated, combining technical safeguards, operational practices, and institutional governance. However, unpredictable capability emergence, limited model interpretability and a persistent 'evaluation gap' complicate risk assessment. This underscores the urgent need for rigorous standards, validated metrics, and harmonised evaluation frameworks to ensure safe and sustainable deployment of general-purpose AI systems (Bengio et al., 2026).

## Conclusion

The development of automated driving has followed a near century-long trajectory, from infrastructure-guided prototypes to vehicle-centric systems and now to data-driven, AI-enabled approaches built upon deep learning. Each technological leap has brought new questions about how safety can be assured, demonstrated, and trusted. As automation systems evolve beyond deterministic rules toward end-to-end deep learning architectures, the imperative for robust, transparent and standardised data collection becomes paramount.

Safety assurance for automated driving can no longer rely solely on accumulated test mileage or comparisons with human crash rates. Instead, it must be grounded in continuous, data-driven evaluation, integrating evidence from simulation, real-world operation, and post-incident analysis. The principal categories of data generated by AVs are operational, perception, prediction, decision-making, and event data. Together they provide a rich foundation for understanding system behaviour and performance. However, the diversity of sensing technologies, software architectures and proprietary data formats across developers creates major challenges for regulators seeking to make fair and consistent assessments of safety.

This is where initiatives such as BSI's Digital Commentary Driving (DCD) framework can play a transformative role. By standardising how automated vehicles describe their perception, prediction, and decision processes without mandating a specific hardware or software architecture, DCD offers a common language for interpreting AV behaviour. It enables developers to explain why a vehicle behaved as it did in any given situation, supports independent audit and comparison, and allows regulators to assess safety performance transparently without compromising intellectual property.

The wider adoption of such approaches could mark a turning point in data-driven safety assurance. Standardisation of data formats, open interfaces and validated simulation environments will be essential to support meaningful oversight and comparison across the global AV ecosystem. Equally, the responsible sharing of safety-critical data among developers, regulators, insurers and researchers will be vital to sustain the public trust on which the future of automated driving depends. Likewise, consumers may need to accept the secure, confidential sharing of data from AVs to support continuous improvement of the service – a feature that is increasingly prevalent in modern, connected vehicles.

Ultimately, the data imperative is not just a technical challenge but a social responsibility. Transparent, interpretable data provides a credible pathway to demonstrate that automated vehicles are developed and deployed safely, improving continuously over time. If achieved, this will allow automated driving to evolve from promise to practice, delivering measurable safety improvements and realising the long promised societal benefits of safe, clean, efficient and accessible road transport.

## Contributors

Dr Nick Reed – Reed Mobility (lead author and Chair, FISITA Automated Driving Expert Group)

Javier Ibanez-Guzman – Ampere (Renault)

Steffen Knoop – Bosch

Olivier Coppin – Brembo

Costi Shami – General Motors

Fabrice Herveleu – UTAC Group



**BOSCH**



**brembo**



**UTAC**  
ADVANCING MOBILITY SINCE 1924

## References

Bel Geddes, N. (1940). *Magic motorways*. Random House.

Bengio Y. et al. (2026). International AI Safety Report 2026. Accessed on February 2026 from: <https://internationalaisafetyreport.org/sites/default/files/2026-02/international-ai-safety-report-2026.pdf>

Bonnefon, J-F., Černý, D., Danaher, J., Devillier, N., Johansson, V., Kovacikova, T., Martens, M., Mladenovic, M.N., Palade, P., Reed, N., Santoni De Sio, F., Tsinoema, S., Wachter, S., Zawieska, K. (2020) *Ethics of Connected and Automated Vehicles Recommendations on road safety, privacy, fairness, explainability and responsibility*. European Commission. Accessed July 2025 from <https://doi.org/10.2777/966923>

Bosch. (2025). ADAS end-to-end AI powered by GPT technology. Accessed November 2025 from [https://www.bosch-mobility.com/media/global/company/events/experience-days/experience-days-north-america/pdf/shared\\_adas\\_e2e\\_ai.pdf](https://www.bosch-mobility.com/media/global/company/events/experience-days/experience-days-north-america/pdf/shared_adas_e2e_ai.pdf)

Cheon, S. (2003). An overview of Automated Highway Systems (AHS) and the social and institutional challenges they face (UCTC Working Paper). University of California Transportation Center. Accessed July 2025 from <https://escholarship.org/uc/item/8j86h0cj>

Di Lillo, L., Gode, T., Zhou, X., Scanlon, J. M., Chen, R., & Victor, T. (2024). Do Autonomous Vehicles Outperform Latest-Generation Human-Driven Vehicles? A Comparison to Waymo's Auto Liability Insurance Claims at 25.3M Miles.

Dickmanns, E. D., & Zapp, A. (1987). A dynamic visual perception system for vehicle guidance. *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI)*, 2, 875–881.

Dickmanns, E. D. (2002). Vision for ground vehicles: History and prospects. *International Journal of Vehicle Autonomous Systems*, 1(1), 1–44. <http://dx.doi.org/10.1504/IJVAS.2002.001804>

Gibson, G. W.: Why Don't We Have... CRASH-PROOF HIGHWAYS. *Mechanix Illustrated*, June 1953, 58–60 and 184 (1953)

Engelking, C. (2017). The 'driverless' car era began more than 90 years ago. *Discover Magazine*. <https://www.discovermagazine.com/technology/the-driverless-car-era-began-more-than-90-years-ago>

Favaro, F., Schnelle, S., Fraade-Blanar, L., Victor, T., Peña, M., Webb, N., Broce, H., Paterson, C. & Smith, D. (2025). Determining Absence of Unreasonable Risk: Approval Guidelines for an Automated Driving System Release. <https://arxiv.org/abs/2505.09880>

Fraade-Blanar, L., Favaro, F. M., Engstrom, J., Cefkin, M., Best R., Lee, J., Victor, T. (2025). Being good (at driving): Characterizing behavioral expectations on automated and human driven vehicles. Waymo LLC. Accessed July 2025 from [https://storage.googleapis.com/waymo-uploads/files/documents/safety/Drivership\\_Being\\_Good\\_at\\_Driving.pdf](https://storage.googleapis.com/waymo-uploads/files/documents/safety/Drivership_Being_Good_at_Driving.pdf)

Frewer, L. J., Scholderer, J., & Bredahl, L. (2003). Communicating about the risks and benefits of genetically modified foods: The mediating role of trust. *Risk Analysis*, 23(6), 1117–1133. <https://doi.org/10.1111/j.0272-4332.2003.00385.x>

- Kalra, N., & Paddock, S. M. (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? (Research Report RR-1478-RC). RAND Corporation. <https://doi.org/10.7249/RR1478>
- Kaur, K., & Rampersad, G. (2018). Trust in driverless cars: A systematic review. *Transportation Research Part F*, 59, 135–149. <https://doi.org/10.1016/j.trf.2018.08.008>
- Kim, Younghwan & Kim, Minki & Kim, Wonjoon, 2013. "Effect of the Fukushima nuclear disaster on global public acceptance of nuclear energy," *Energy Policy*, Elsevier, vol. 61(C), pages 822-828. <https://doi.org/10.1016/j.enpol.2013.06.107>
- Korosec, K. (2020). Uber in talks to sell ATG self-driving unit to Aurora. TechCrunch. Accessed July 2025 from <https://techcrunch.com/2020/11/13/uber-in-talks-to-sell-atg-self-driving-unit-to-aurora/>
- Korosec, K. (2023). Cruise pauses all driverless robotaxi operations to 'rebuild public trust'. TechCrunch. Accessed July 2025 from <https://techcrunch.com/2023/10/26/cruise-pauses-all-driverless-robotaxi-operations-to-rebuild-public-trust/>
- Korosec, K. (2024). RIP, Cruise robotaxi. TechCrunch. Accessed July 2025 from <https://techcrunch.com/2024/12/12/rip-cruise-robotaxi/>
- Kusano, K.D., Scanlon, J. M., Chen, Y., McMurry, T. L., Gode, T., Victor, T. (2025). Comparison of Waymo Rider-Only Crash Rates by Crash Type to Human Benchmarks at 56.7 Million Miles. *Traffic Injury Prevention*, 26(5), 608–621. <https://doi.org/10.1080/15389588.2025.2499887>
- Kyriakidis, M., Happee, R., & de Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire. *Transportation Research Part F*, 32, 127–140. <https://doi.org/10.1016/j.trf.2015.04.014>
- Leahy, K., Asgari, H., Dennis, LA., Feather, MS., Fisher, M., Ibanez-Guzman, J., Logan, B., Olszewska, J., Redfield, S., (2026). Open-world verification: a grand challenge for autonomous systems. *Proceedings of the IEEE*, in print.
- Parker, C. (2023). More British troops killed at the wheel than in war. *The Times*. Accessed July 2025 from <https://www.thetimes.com/uk/politics/article/more-british-troops-killed-at-the-wheel-than-in-war-q6g82vdf>
- Pomerleau, D. A. (1989). ALVINN: An autonomous land vehicle in a neural network (Technical Report CMU-CS-89-107). Carnegie Mellon University, School of Computer Science.
- Reed, N., Balcombe, B., Spence, P., Khastgir, S., & Fleming, N. (2023). A review of CAV safety benchmarking and a proposal for a "Digital Commentary Driving" technique. British Standards Institution and the Centre for Connected and Autonomous Vehicles, June 2021.
- Reed, N., Houghton, E., Pressley, A., Vigilia, K., & Dent, D. (2023). Self-Driving But Guided By People – How to make automated vehicles ethical. Reed Mobility report for the Rees Jeffreys Road Fund. Accessed November 2025 from <https://shorturl.at/CFwz4>
- Tesla. (2025). Tesla Vehicle Safety Report. Accessed November 2025 from [https://www.tesla.com/en\\_gb/VehicleSafetyReport](https://www.tesla.com/en_gb/VehicleSafetyReport)

UNECE 155 (2021), UN Regulation No. 155 – Cyber security and cyber security management system, E/ECE/TRANS/505/Rev.3/Add.154

UNECE 156 (2021), UN Regulation No. 156 – Software update and software update management system,

UNECE 157 (2021) UN Regulation No. 157 – Automated Lane Keeping Systems (ALKS,) E/ECE/TRANS/505/Rev.3/Add.156  
United Kingdom. (2024). Automated Vehicles Act 2024 (c. 10). Accessed November 2025 from <https://www.legislation.gov.uk/ukpga/2024/10>

Wayve. (2025). Driving the World, Sight Unseen: Wayve's AI-500 Roadshow Hits 90 Cities in 90 Days. Accessed November 2025 from <https://wayve.ai/thinking/ai-500-roadshow-90-cities/>

Wetmore, J. M. (2003). Driving the dream: The history and motivations behind 60 years of automated highway systems in America. *Automotive History Review*, (7), 4–19.