Assurance Through Safety Cases—There’s a Claim For That

By Nat Beuse and Chan Lieu
Aurora Innovations, Inc.

Aurora Submission for ESV 2023

Executive Summary

Designing, developing, testing, and deploying an Automated Driving System (ADS) for use on public roadways in the United States is challenging for a variety of reasons, including for the ADS developer in defining and describing their approach for ensuring the safety of their vehicles. An autonomous vehicle is subject to National Highway Traffic Safety Administration (NHTSA) motor vehicle safety requirements, despite there being no defined Federal Motor Vehicle Safety Standards (FMVSS) that govern ADS performance requirements. The operation itself may be subject to other federal safety, state, and local laws and regulations depending on the type of operation (e.g., commercial motor vehicle or passenger service operation) and operating location. In addition, there is federal voluntary guidance containing priority safety design elements and a growing number of relevant industry-developed consensus standards and best practices available to an ADS developer to consider and incorporate in the design of their ADS. In navigating these various regulatory frameworks, standards, and best practices, the ADS developer is still ultimately responsible for defining and ensuring safety for their own vehicles. A safety case-based approach is a valuable way to provide such assurance.

A safety case is a structured argument, supported by evidence, intended to justify that a system is acceptably safe for a specific application in a specific operating environment. While this approach is not entirely new – safety cases have been incorporated into other safety-critical industries – safety cases for the development of autonomous vehicles are novel.

A safety case-based approach creates value through both flexibility and a high degree of rigor, if applied correctly. It is flexible because it provides the ADS developer with the latitude to determine what claim to make, and it is rigorous because there must be evidence to substantiate it. For example, there are now several publicly available voluntary industry standards and guidance spanning many important topics related to the development and safe operation of an ADS. These topics include functional safety, behavioral safety, and safety assurance for machine learning systems.[1] The emergence of these standards provide varying perspectives that ADS developers should consider and how an ADS developer implements these standards can be the basis of a safety case claim related to adhering to industry standards.

This paper will present Aurora's experience and lessons learned in developing and implementing its own Safety Case Framework. This includes discussion regarding how Aurora integrates existing industry standards into the ADS development process, while also

[1] https://safeautonomy.blogspot.com/2022/04/maturity-levels-for-autonomous-vehicle.html
https://www.eetimes.com/ul-4600-draft-puts-safety-onus-on-av-hopefuls/

building on them by incorporating into the development process vehicle product engineering requirements, enterprise wide processes, and operational elements (such as a Safety Management System). A safety case-based approach is important to ensure that the integration of many new, overlapping standards is managed correctly. And ultimately, a safety case-based approach provides transparency and insight into safety assurance.

What is a Safety Case?

A safety case is a structured argument, supported by evidence, to justify that a system is acceptably safe for a specific application in a specific operating environment.[2] A structured argument includes a specific claim – in Aurora's case, that our self-driving vehicles are acceptably safe to operate on public roads – that is then divided into subclaims. These subclaims may be further broken down into additional subclaims that ultimately result in evidence to substantiate the claim.

Safety cases are not a new concept – they have been widely used in various other safety-critical industries, such as oil and gas exploration, aviation, rail, medical devices, and nuclear energy.[3] (See Fig. 1 for other examples and timelines)

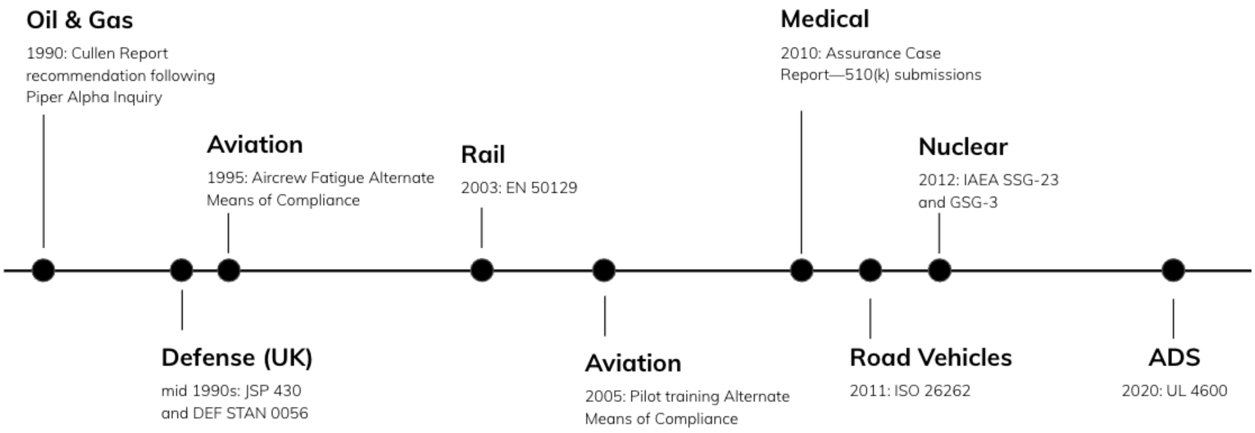
Oil & Gas
1990: Cullen Report recommendation following Piper Alpha Inquiry
Defense (UK)
mid 1990s: JSP 430 and DEF STAN 0056
Aviation
1995: Aircrew Fatigue Alternate Means of Compliance
Rail
2003: EN 50129
Aviation
2005: Pilot training Alternate Means of Compliance
Medical
2010: Assurance Case Report—510(k) submissions
Road Vehicles
2011: ISO 26262
Nuclear
2012: IAEA SSG-23 and GSG-3
ADS
2020: UL 4600

Fig. 1 Safety cases in other industries

Aurora has adopted a safety case-based approach because we believe it is the most logical and efficient manner to show and explain how Aurora determines that our self-driving vehicles are acceptably safe to operate on public roads. The heart of our safety case is a

[2] Defence Standard 00-56 Issue 4 (Part 1): Safety Management Requirements for Defence Systems. UK Ministry of Defence.
[3] See Cullen, W. Douglas. *The Public Inquiry into the Piper Alpha Disaster*. London, UK, 1990.

structured argument, supported by evidence, to demonstrate the claim for why our vehicles are acceptably safe for operation on public roads.

No single piece of evidence captures the totality of safety. There are complex interactions and relationships between the many elements that go into developing an ADS and it can be difficult to track and trace these interactions. We can provide various pieces of evidence, but without the appropriate context it would be difficult to understand why and how these pieces of evidence are relevant.

Ultimately, evidence without a claim is simply trivia and, conversely, a claim without evidence is baseless. A safety case-based approach brings these two essential concepts together in a logical manner to effectively show the work that we have undertaken to determine our vehicles are acceptably safe to operate on public roads.

Safety Case Framework vs. Safety Cases

Aurora's Safety Case Framework is designed to be adaptable to different vehicle platforms and operational contexts, and is the superset that captures all the claims that Aurora argues are necessary to safely deploy our ADS. It provides a logical mechanism to describe, characterize, and justify that the ADS is acceptably safe to operate. The structured argument of the Safety Case Framework outlines ***what*** the completed safety argument must achieve without providing prescriptive requirements on ***how*** it must be achieved (see Fig. 2). This abstraction is intentional and provides flexibility for different engineering teams to define and develop the evidence to substantiate the claim. Aurora's safety cases are derived from this framework and each is simply the compilation of the evidence addressing the relevant claims.

Aurora's Safety Case Framework is built upon five principles that describe our approach to developing our self-driving technology – Proficient, Fail-Safe, Continuously Improving, Resilient, and Trustworthy. The Principles are the broad categories that guide and describe our goals for developing safe autonomous vehicles, including:

- **Proficient**: The vehicle is acceptably safe during normal driving. Essentially, everything is working as intended.

- **Fail-Safe**: The autonomous vehicle is acceptably safe when there is a fault or failure. We design our vehicles in such a way that, if some component fails (like if a sensor is damaged or a tire blows out), the vehicle should behave in a manner that does not endanger its passengers or other road users.

- **Continuously Improving**: Aurora is committed to continuously improving. We are constantly learning and striving to identify, evaluate, and resolve anomalies that could affect the safety of the vehicle.

- **Resilient**: Our vehicles are acceptably safe in the case of reasonably foreseeable misuse and unavoidable events. For example, our cybersecurity-related claims mostly reside under this principle.

- **Trustworthy**: The public can have confidence in not only Aurora's autonomous vehicles, but our entire company – that we not only design, build, and test our self-driving vehicles in a dependable manner, but also that we have a safety and organizational culture in place to quickly address and resolve issues.

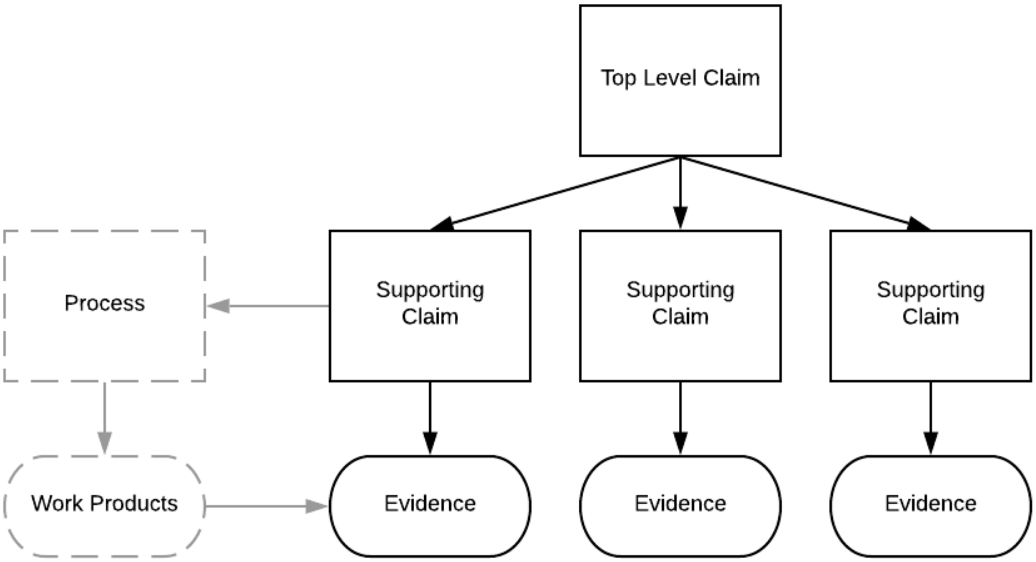

Fig. 2 Top down claim structure

Safety Case Tailoring

We use the Safety Case Framework to create a tailored safety case, taking care to define its specific context and application in each instance. A tailored safety case is a subset of the overall framework that includes the claims and subclaims appropriate for a specific use case, such as autonomy operations with a vehicle operator (VO). Aurora has multiple tailored safety cases, the primary two of which are centered around—[4]

1) operating in autonomy on public roads with VOs onboard (VO Road), and

[4] Aurora also has other safety cases based on specific hardware configurations (e.g. sensors and computing hardware) on various vehicle platforms (e.g. Class 8 tractor or passenger vehicle).

2) operating in autonomy with no VOs onboard (NVO Commercial).

Since the VO Road safety case is designed for ADS on-road operations with a VO present in the vehicle, the claims in this tailored safety case emphasize VO hiring, training, and certification and recertification, as well as driver monitoring systems. However, claims related to VO hiring, training, and certification and recertification become irrelevant when we progress to the NVO Commercial safety case. As a result, those VO related claims are no longer applicable, or are "tailored out." At the same time, we "tailor in" other claims, such as the ADS's ability to execute a minimum risk maneuver to achieve a minimum risk condition when a fault is detected, for the NVO Commercial safety case since there is no longer a VO in the vehicle to intervene.

This tailoring of the framework is unique and depends on the assumptions, claims, and arguments we make in each safety case. As a result, while the Safety Case Framework may be the superset, the tailoring activity is essential to guide evidence development. Aurora already maintains VO safety cases and expects to maintain multiple safety cases for NVO and different vehicle platforms (e.g., Class 8 trucks vs. passenger vehicles) so that all vehicle configurations we operate on public roads have an associated safety case.

Types of Claims

Aurora's Safety Case Framework claims can be broadly categorized into three major buckets: product, process, and operational safety. Product claims are related to the engineering of the ADS, and the evidence to support those claims will change over time as we iterate and improve the ADS's capabilities. Process claims are related to our process for how we develop the ADS and the evidence supporting those claims tend to be less frequently updated. Since most companies will look to improve their development processes, we are not implying that the evidence for this will never change, but it will do so less frequently. Finally, operational claims are related to how the ADS is operated on public roads.

Aurora's Safety Case Framework is more than a set of product requirements by encapsulating contributions from the entire company. This means that, in addition to the engineering involved with developing the ADS, we are incorporating non-engineering contributions from other portions of the company, such as government relations or legal, that are critical to deploying an ADS. By aligning cross-functional contributions when developing and deploying the ADS, we have created a safety culture throughout the company in which everyone at Aurora plays a role in safety, and our safety case helps recognize all of those contributions in a manner that requirements alone don't capture.

Incorporating Standards & Best Practices

The National Highway Safety Administration (NHTSA) has promulgated a number of FMVSS that define performance standards for automotive safety requirements that apply to all motor vehicles, including autonomous vehicles. These standards include clear test procedures, parameters, and pass/fail criteria that motor vehicle manufacturers meet when designing their vehicles. While NHTSA has not yet published FMVSS specifically related to ADS performance requirements, many ADS developers leverage voluntary industry developed standards and best practices as they design, develop, and deploy their ADS.

The industry, through standards development organizations, has produced a number of useful voluntary standards and best practices for ADS developers to consider and adopt when designing their ADS. These voluntary standards and best practices cover a variety of safety relevant topics, including system and operational safety and ADS verification and validation, operational design domain, and cybersecurity. Notable standards and best practices ADS developers can consider include ISO 26262 Road vehicles – Functional Safety [5] and ISO 21448 Road vehicles – Safety of The Intended Function (SOTIF)[6] regarding system safety and a number of Automated Vehicle Safety Consortium (AVSC) best practices, ranging from ODD definitions, to first responder interactions, to metrics and methods for assessing safety outcomes (see Table 1).

[5] ISO 20262 Road vehicles – Functional safety, 2018. https://www.iso.org/standard/68383.html
[6] ISO 21448 Road vehicles – Safety of The Intended Function, 2019. https://www.iso.org/standard/70939.html

AVSC ID	Title
AVSC0006202103	Metrics and Methods for Assessing Safety Outcomes of Automated Driving System (ADS)
AVSC00009202208	Interactions Between ADS-DVs and Vulnerable Road Users (VRUs)
AVSC0007202107	Information Report for Adapting a Safety Management System (SMS) for Automated Driving System (ADS) SAE Level 4 and 5 Testing and Evaluation
AVSC0005202012	First Responder Interactions with Fleet-Managed Automated Driving System-Dedicated Vehicles
AVSC00008202111	Evaluation of Behavioral Competencies for Automated Driving System Dedicated Vehicles (ADS-DV)
AVSC0004202009	Data Collection for Automated Driving System Dedicated Vehicles to Support Event Analysis
AVSC00002202004	Describing an Operational Design Domain: Conceptual Framework and Lexicon
AVSC00001201911	Safety operator selection, training, and oversight procedures for automated vehicles under test

Table 1 AVSC best practices

ADS developers each have their own processes to evaluate whether and how to adapt and conform with these various standards and best practices. That type of work is captured in Aurora's Safety Case Framework in multiple different places. For example, under one Aurora claim –"G5 The Self-Driving Enterprise is trustworthy" – there is a subclaim related to how prevailing industry best practices and standards are reviewed and adherence documented (G5.1.1.1.3). Evidence to support this subclaim might include a process of how Aurora periodically surveys the publication of new industry standards or best practices. It might also include a library of identified standards and the documentation on how Aurora is conforming to these standards. We would also expect to also see evidence in the G1 Proficient principle related to how these standards or best practices are traced to requirements. Tying all these different pieces of evidence together is what substantiates the claim.

Safety Case Challenges

Maintaining flexibility

Aurora chose to be less prescriptive in developing the safety argumentation and the claim language in its Safety Case Framework because flexibility is key to our G3 Continuously Improving principle. This flexibility enables broader applicability and longevity before having to revisit and update the overall framework. For example, in our claim regarding incorporating industry standards and best practices described above, the claim language itself states that "prevailing industry best practices and standards are reviewed and adherence documented, on a continual basis." It purposely does not specify *which* industry best practices and standards because the identification and selection of those standards is part of the point of the claim. This flexibility enables us to retire a best practice or standard should it become outdated or deprecated, without having to redefine the argument. Furthermore, prescriptive language has the risk of introducing rigidity and risk the argument becoming invalid.

Claim ownership

When it comes to project management of a Safety Case Framework claim, it is important to identify a single claim owner who will be designated as responsible for ensuring that the evidence to support the claim is completed. While this may be straightforward in a hierarchical structure, it becomes more difficult to manage with cross functional teams of an organization. Cross functional teams bring together resources from multiple parts of the organization to jointly work together (e.g., incorporating resources from product development, software, and hardware). However, the challenge now becomes which functional organization would be responsible for overseeing work on the single claim. Ultimately, as with many development efforts, coordination and communication is key to ensuring that all collaborators understand the task and are aligned on the deliverables.

Adapting UL 4600

UL 4600 is intended to help ensure that an acceptably thorough consideration of safety for an autonomous product has been performed during the design process and will continue to be done throughout the system life cycle. It does so by emphasizing repeatable assessment of the thoroughness of a safety case.[7] It is also the first standard to address the entirety of safety assurance in the design, development, and deployment of automated vehicles.

[7] UL Standard for Safety for Evaluation of Autonomous Products, UL 4600, 2nd Edition, March, 2022.

While UL 4600 was comprehensive in many aspects of autonomous vehicle development, there were several gaps that were deliberately out of scope by the authors, most notably the road testing of prototype vehicles that include human operators responsible for supervising the autonomous systems.[8] Since Aurora is currently testing vehicles in development with VO supervision of the ADS, it was imperative to incorporate those safety arguments into our overall Safety Case Framework. By developing those additional claims, we are then able to control when they come into scope.

Communicating a safety case

Since safety cases are new for the automotive industry, it is necessary to educate industry stakeholders, regulators, and others on what safety cases are, how they are constructed, and how they can be interpreted. This initial effort is necessary because safety cases can be complex and difficult to comprehend, often requiring context and explanation in order to fully grasp the safety argument. Beyond this initial education, there is also the challenge of bringing all the different pieces of evidence together in order to tell that story. A comprehensive report that breaks down the safety case argument, details the claims, and puts the various pieces of evidence into the necessary context, without forcing the reader to wade through potentially hundreds or thousands of pieces of evidence, would be the most effective way to summarize that work.

Conclusion

We've built our Safety Case Framework and each of its five supporting principles to guide responsible ADS development. Each principle is supported by multiple claims and will be substantiated by hundreds of pieces of evidence. Only by validating our system with hardened evidence through this process can we build confidence in the Aurora Driver's ability to safely operate on public roads without a human driver.

We believe this is a powerful tool that not only can guide a company as it develops ADS technology, but can also be useful in telling a coherent safety story. At the same time, it would be inadvisable to mandate a safety case because doing so runs the risk of turning this introspective activity into a check-the-box exercise whereby the motivation would be rooted in compliance and whose value would be further diminished via enforcement action. We believe every company building this transformational technology should openly share their safety case. At Aurora, our safety case shows that we are doing more than just committing to safety in principle, we are putting safety into practice.

[8] *Id.* section 2.1.2