Automotive AI: Help or Hindrance?



In this article Dr Helen Monkhouse discusses the impact that introducing Artificial Intelligence (AI) into automotive products might have. Seen as an enabling technology for highly automated driving and driver support systems, AI also poses a disruptive challenge to vehicle safety assurance, so will it turn out to be a help or hindrance to the continued safe development of vehicle systems?

Travelling in an early 20th Century motor vehicle was a dangerous pursuit, and with vehicles being heavy and difficult to control, serious accidents were common. Since those early days road safety and vehicle safety have both seen a gradual evolution. During the early 1900's windscreen wipers, rear view mirrors and laminated windscreens were fitted to vehicles. The 1950's saw the introduction of 'passive' safety systems, including seat belts and the first airbags, that sought to minimise accident severity. 'Active' safety systems, like Anti-lock Braking Systems (ABS) and Electronic Stability Control (ESC), that support the driver to prevent a hazardous situation becoming an accident, are now standard.

The complexity of vehicle safety features continues to grow, with Advanced Driver Assistance Systems (ADAS) providing 'tactical' safety to the driver. That is, preventing driving situations becoming critical through the introduction of features, such as Automatic Emergency Braking (AEB), blind-spot monitoring and Lane Keeping Assistance (LKA). By their nature, 'tactical' safety systems require a comprehension of the vehicle's operating environment, or proper situational awareness, to operate correctly. It is for ADAS, Automated Driving (AD), and ultimately driver-less vehicles, where AI is seen as the enabler that will help propel these vehicle features into series production.

Although AI, and more specifically machine learning (ML), enables sophisticated vehicle perception systems to be realised, as a technology it possesses properties that differ from those technologies upon which existing safety practices are based. This means that AI technology is potentially a hindrance to the continued safe development of vehicle systems.

To unpick this help or hindrance question, I reflect on AI's use in automotive products, before then considering how this impacts product safety assurance. To keep article size manageable, I am ignoring the use of AI in automotive engineering development tools. Having considered how AI might impact automotive product safety assurance, I then reflect on what aspects of the classical design paradigm still have utility when AI technologies are deployed into safety relevant product. I conclude by returning to the question – Automotive AI: help or hindrance?

What is AI and where is it being used?

Artificial Intelligence

The International Standard on Artificial Intelligence (AI) ISO/IEC 22989:2022 defines an AI system as an "engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives." The standard also notes that the "engineered system can use various techniques and approaches…to develop a model to represent data, knowledge, processes, etc. which can be used to conduct tasks," and that AI can "operate with varying levels of automation" [1]. The creation of such a rational agent is concerned with developing machines capable of achieving goals and deliver the best outcome in the presence of uncertainty.

The theoretical techniques that underpin AI are not new. If we think about control theory, business operations or statistics, then it is typical to minimise or maximise to some function. It is high performance computing that now makes the use of such theoretical techniques tenable for industries like automotive. However, a potential limitation of an objectives led approach is the need to fully define the objectives [2]. For a self-driving vehicle, superficially the objective might be to move the passengers safely from Point A to Point B. However, when considered against external factors like, errors by other road users, bad weather or component failures, achieving strict adherence to this safety goal may well result in the AI based route planning function choosing never to route the vehicle out of its garage!

"Achieving strict adherence to this safety goal may well result in the AI based route planning function choosing never to route the vehicle out of its garage!"

AI use in automotive applications

As mentioned in the introduction, it is ADAS and AD perception systems where AI is being deployed within safety relevant automotive product. Side-stepping the issue of safe AI, by developing traditional procedural-based vehicle perception systems has limited utility. By simply considering the data representation of pedestrian forms in an Operational Design Domain (ODD), one can see that to fully define a pedestrian, such that it can be recognised by a procedural algorithm, becomes an intractable task.

The introduction of AD vehicle features that allow the driver to relinquish control of the

driving task to the automation, for a time, introduces a requirement for driver monitoring. Typically, camera-based systems within the passenger compartment monitor the driver's eye gaze and eye opening to assess alertness and fatigue. Decisions regarding whether to continue in AD mode, or when to hand-back control to the driver may be predicated on the assessment of the driver's current attentiveness. For example, conditional driving automation systems, like BMW's Personal Pilot or Mercedes Benz's Drive Pilot, allow the driver to potentially engage in non-driving task for several seconds while the feature is operational.

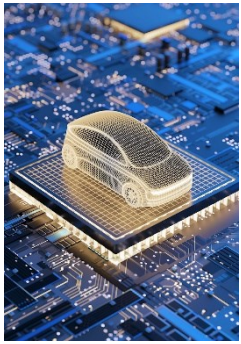
Perhaps not obvious to the end user, but AI based techniques are being employed in control adaptation too. Battery management systems that use adaptation techniques to improve the performance and longevity of electric vehicle batteries are in development, and adaptation techniques that improve system robustness, are widely used in engine and transmission control. Although typically deployed in the main control path, and not as part of the safety mechanism, the feasibility of safety monitoring adaptation is also being explored [3].

Automotive safety assurance

Although a complex system of systems, many of the engineering principles that underpin the development of the modern passenger vehicle are based on classical mechanical engineering principles. In this classical design paradigm, the vehicle comprises several discrete independent systems (what ISO 26262 refers to as an Item [4]). These Items deliver vehicle functions like power steering control, braking and stability control, that exhibit predictable behaviour that remains consistent throughout the vehicle's life. Today's new-vehicle homologation process is predicated on this property. That is, having verified that a new vehicle type meets the regulatory performance targets, assuming manufacturing quality is controlled and maintained, then the performance of the millionth vehicle produced will still adhere to those same regulatory targets.

Items like those described above will typically comprise an electronic control unit, having several sensor inputs and one or more output actuators. Requirements can be written that fully describe 'what' a given Item needs to do, from which the design specification can be created detailing 'how' the system implements its requirements. Assuming the Item's intended behaviour (the 'what') is not in itself unsafe, then functional safety is achieved if the architectural design prevents random hardware failures causing the Item's actuators to lead to harm, and if sufficient rigour has been applied during development to avoid systematic design errors.

Achieving safety of the intended function (SOTIF) becomes more involved for systems that sense their environment. The International Standard ISO 21448 describes the process of achieving SOTIF, in which system limitations (called functional insufficiencies) and the conditions under which these limitations may manifest (called triggering conditions) are identified, understood and managed [5]. A difficult task made more difficult when the system contains an AI component, because the way in which the AI model learns from the data is at a different level to the inferences made by the human analyst. That is, humans will typically analyse the system semantically, considering scene attributes like lighting conditions and the classification and position of objects, while the AI model operates on the underlying data.



The classic design paradigm assumes that components experience binary failures. That is, the fault is either present and remains present or it is not. For hardware components, the component tolerance could drift, or an open / short-circuit fault could occur. This will typically be because of aging and follow an exponential distribution. Software component failures are not caused by aging. Instead, a software systematic fault (i.e., software 'bug') can lay undiscovered, but will manifest itself every time the input conditions are such to generate the failure. If we think of an AI component as being an executable element (see Figure 1) then an AI component can also experience failures due to hardware random aging and software systematic failures, because of the hardware and software platforms on which the AI model runs. However, in addition to the classic random and systematic failures, AI components can exhibit misclassification errors.



For example, a ML vision system failing to classify an object as a pedestrian. However, given the relative complexity of the input space, the shape of the failure probability distribution is unknown. Potential methods to address these uniquely AI related faults are discussed below.

'Tactical' safety systems, whose correct and safe behaviour is predicated on correctly perceiving the vehicle's surroundings, challenge the classical design paradigm. Not only does fully defining the 'what' under all conceivable operating conditions become problematic, but the use of AI techniques like Deep Neural Networks (DNNs), makes understanding 'how' the system achieves the 'what' less explainable too.

Is the automotive product safety process broken?

It would be easy to dismiss the use of AI components in safety relevant applications, but this could be to the detriment of the potential benefits that deploying the technology might bring. To entertain using AI in a safety related application requires understanding the constituent parts of the AI component, to assess the impact that each might have on system safety, and in doing so devise means by which to mitigate potential errors.

Before delving into the impact that an AI component might have on safety, it is important to be explicit about what an AI component comprises. For the purposes of this discussion let us assume that to deploy an AI model, like a DNN, as executable code into an automotive application we need to create an Embedded AI System.

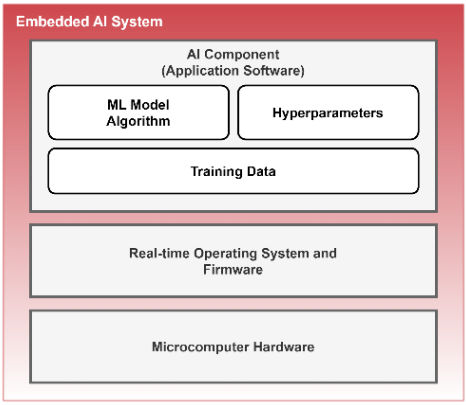


Figure 1 Embedded AI System

The Embedded AI System architecture is depicted in Figure 1. For the purposes of this discussion the architecture includes application software, an operating system, hardware specific firmware and computer hardware. Of relevance here is the application software, or AI component, which comprises

the model, hyperparameters[1], and training data. Supporting processes, not included on the diagram, but pertinent to the discussion include data management, the training approach (e.g., supervised learning, reinforcement learning) and test strategies.

How can an AI model cause harm?

At a superficial level the propagation of an AI model fault, through to unsafe vehicle behaviour can be represented by the diagram in Figure 2. Depending on its origins, the unsafe vehicle behaviour experienced can be classified as malfunctioning behaviour [4], hazardous behaviour [5], or loss of controllability due to misuse.[2] Although these classifications are unhelpful to the vehicle users experiencing the unsafe vehicle behaviour, they do illustrate how the different automotive standards dovetail together to address potential safety relevant AI faults.



Figure 2 AI component error propagation

From a safety perspective, techniques to prevent systematic faults in source code (e.g., divide by zero, incorrect variable names) and random hardware faults (e.g., component open or short-circuit) are addressed by the automotive functional safety standard ISO 26262:2018 [4]. ISO PAS 8800, a new international standard in development, seeks to address sources of functional insufficiencies in the AI technology capable of affecting vehicle safety [6].

"An AI component can be only as good as the model and the data that creates it"

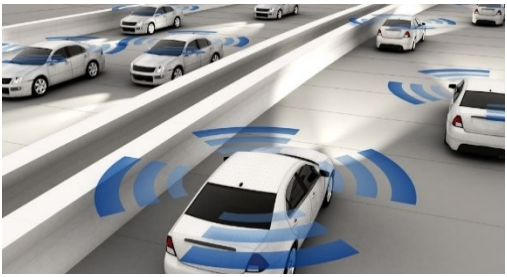
An AI component can be only as good as the model and the data that creates it. Errors in the specification, collection, generation and classification of data used to train and test the AI model can all result in functional insufficiencies, which can then contribute to a vehicle system error. Training data having missing data sets can lead to the misclassification of objects in the field-of-view. That is, semantic properties of the intended ODD may not be represented in the training data. In addition, the distribution of training data might not give a good approximation of the real-world characteristics, or it may shift over time. For example, if new car type or colour is introduced after the AI model has been trained and introduced into the field. The brittleness of DNNs also has the potential to impact safe behaviour. Common input

[1] Hyperparameters are the external parameters used to specify the AI model architecture. For example, the number of nodes and layers in a neural network, or the number of branches in a decision tree.

[2] Specifically, ISO 21448 refers to this type of misuse as Reasonably Foreseeable Indirect Misuse (RFIM) [4]. Examples of which could include, the driver falling asleep or unbuckling their seat belt and climbing into a passenger seat while an AD system is responsible for vehicle control.

perturbations like signal noise, weather conditions and objects appearing in atypical orientations all have the potential to produce output misclassification [7]. Even with correctly distributed data, a well-documented training error for DNNs is 'overfitting'. This can cause a DNN to give a corner case a wrong high-confidence classification.

In traditional system safety development lifecycles, Validation and Verification (V&V) both play an important role in building confidence in the system prior to deployment. When referring to V&V we perhaps immediately think of physical testing, however a significant part of the verification activity is system verification by analysis. That is, verifying that both the system architecture delivers 'what' is specified by the safety requirements, but also 'how' the safety architecture implements the safety requirements does not introduce new mechanisms for harm. A verification challenge for systems deploying AI models is the potential complexity of the model's architecture. This presents a scalability problem, as it is not uncommon for the number of parameters within a DNN to be into the millions.



It is a widely held view that for a system to be safe it must also be secure. Before passenger vehicles became networked, system security was a matter of physically preventing access to sensitive components, for example, using anti-tamper screws. High profile security breaches [8] have demonstrated the potential vulnerabilities that are introduced once vehicles are networked. However, by following cybersecurity engineering best practice [9] these types of vulnerabilities can be managed. The situation becomes more complex with the introduction of AI. Adversaries no longer need to gain access to the system in order to exploit that system's vulnerabilities. Instead, an adversary attack can exploit system vulnerabilities simply by manipulating the vehicle's environment. For example, by applying stickers to objects in the environment, or by partially obscuring a vehicle sensor.

The techniques that remain applicable

Irrespective of the technology, the underlining philosophy of safety engineering is to understand the causes of harm and then to devise mechanisms to prevent, avoid or mitigate that harm. What does change are the conditions under which risk should be subjectively assessed and therefore the integrity to which the product should be engineered. The existing automotive risk model [4] assumes that the driver is integral to vehicle control and ultimately responsible for vehicle safety. Risk mitigation under this paradigm involves deploying safety mechanisms that either reduce accident severity or increase the driver's controllability to a point where any residual risk would be rated as ASIL QM[3] [4]. For AD, part of the SOTIF process involves defining an acceptable residual risk [5]. For example, the number of persistent false negative pedestrian misdetections permissible by a DNN in a 30 frames per second video sequence. Defining acceptance criteria for an AD vehicle feature is both non-

[3] The ISO 26262 risk model classifies risk by considering accident severity, the probability of the hazardous event, and the driver's ability to control the hazardous event. The lowest level of risk is defined as QM – a level of risk where applying quality management principles are sufficient to achieve safety.

trivial and not uniquely an AI problem. As AD feature familiarity, regulation, and societal acceptance matures over time then best practice criteria will likely develop for the different features introduced.

The ability to shut-down a faulty system, reliability concerns and cost sensitivity of automotive production has meant that redundant safety architectures, such as a 2 out of 3 voting systems, are uncommon. However, AD availability requirements are increasing the need for redundancy. As are the introduction ML perception systems, where inputs from various sensor technologies can be fused together to improve overall system robustness.

Understanding how faults propagate and contribute to vehicle level hazardous behaviour, and ultimately an accident, remains important for systems that include AI components, as does the context in which the hazardous behaviour occurs. For example, the safety consequences of an AEB perception system's failure to detect a pedestrian far from the vehicle is less than the failure to detect a pedestrian located directly in front of the vehicle. In addition, plausibility checks can help the system to gain confidence in a prediction. For example, it is unlikely that a genuine object will disappear from the field-of-view within a few milliseconds. Instead, one would expect to see a real object (e.g., a pedestrian) remain in a video image for multiple frames.

I have long felt the Functional Safety Concept (FSC) to be the most important ISO 26262 work product, because, if written well, it conveys a logical description of how the item achieves safety. As a minimum the FSC should contain a description of the preliminary architecture (logical), the safety mechanisms, and the warning and degradation strategy. For a system containing an AI component, conveying the concepts used to achieve safety is paramount, as safety concepts will likely embody complex acceptance criteria and plausibility considerations. For high integrity systems ISO 26262 recommends, modelling the FSC in a language such as SysML [4]. This remains true for systems containing AI components. However, given the potential complexities and subtleties involved I feel that when the system contains an AI component, then prose should accompany the SysML model, to provide a written narrative of the concepts deployed.

Verification and Validation

In the 'How can an AI model cause harm?' section above, I discuss the importance of verification by analysis, both for achieving functional safety and SOTIF, but suggest that the introduction of AI components makes verification by analysis difficult if not untenable. That said, system level verification by analysis and indeed semantic analysis has an important role to play in AI safety, and that is regarding robustness. Using guideword led analysis (e.g., HAZOP), semantic factors having the potential to affect the sensor system (e.g., windscreen smudges, dirt, blur) can be considered systematically. Once identified, these artefacts can be included into test data, thus improving data completeness. Confidence in the AI model's robustness can be further improved by using classic design of experiment techniques to test input sensitivity. That is, exploring the AI model's sensitivity to the input space by changing input magnitude ranges or by using different combinations of inputs [6]. However, devising a sensitivity measure is a non-trivial task.

Validation remains a challenge for systems containing AI components, particularly testing the unknown or the unpredictable, given that is not possible to extrapolate from known to

unknown test points. For systems containing an AI component, validation refers to the activity of checking that the AI system implements the system safety requirements, therefore it is typically done at vehicle level by the vehicle integrator (e.g., vehicle manufacturer). The focus for the safety validation is achieving full test coverage through the systematic exploration of all relevant test scenarios, and automotive standards now exist dedicated to scenario-based testing [10].

Monitoring vehicle in-field performance as part of a Cybersecurity Management System (CSMS) is now a requirement [9, 11]. Broadening the scope of in-field monitoring to include the identification of new or useful data, can help maintain an AI component's performance through runtime learning. By detecting and sending back out-of-distribution data to a central data centre, the level of uncertainty (i.e., the degree to which the data is out-of-distribution) can be assessed, and ultimately new data sets can be derived and new AI models created. These new AI models can then be put through the normal development cycle and distributed to the vehicles through over the air updates. Allowing in-the-field self-learning may be the utopia, however, from a validation perspective self-learning is considered too problematic and therefore infeasible in the short to medium term.

"Allowing in-the-field self-learning may be the utopia, however, from a validation perspective self-learning is considered too problematic and therefore infeasible in the short to medium term."

In conclusion – help or hindrance?

As I have discussed, a prerequisite for today's 'tactical' vehicle safety features is their ability to obtain and maintain situation awareness. However, given the complexity of the ODD, the ability to define the design space such that traditional procedural algorithms can be used for vehicle perception systems becomes an intractable problem. By virtue of their capability to solve problems (e.g., pedestrian detection) which cannot be fully specified (e.g., characterising all pedestrians within the ODD), deploying AI based systems helps solve the system perception problem.

As we have seen, the introduction of AI systems into automotive product clearly challenges the status quo, by making many of the assumptions that underpin the achievement of functional safety and SOTIF obsolete. Therefore, we could jump to the conclusion that introducing AI components into automotive product is a hindrance to the continued achievement of vehicle safety. However, as this discussion and standards like ISO PAS 8800 highlight, many of the system safety practitioner's existing tools and techniques remain relevant. With less safety requirements typically being defined for the AI component than for a traditional item, the focus of product development effort will need to change. For example, focusing effort into devising test strategies, capable of demonstrating AI model robustness and safety.

As a closing thought, I believe there is a credible route to series production for AI based systems. For me that involves articulating the safety concepts deployed clearly and coherently within the FSC and conveying the safety confidence (with uncertainty) achieved through a comprehensive safety case argument.

References

[1] International Organization for Standardization (ISO), "BS EN ISO/IEC 22989 Information technology – Artificial intelligence – Artificial intelligence concepts and terminology," 2023.

[2] S. Russell, P Norvig, "Artificial Intelligence, A Modern Approach" Pearson, 2022, Fourth Global Edition, ISBN 1-292-40113-3.

[3] F. Botes, et al., "Development and Optimisation of an Adaptive Safety Monitor," SAE WCX 2018.

[4] International Organization for Standardization (ISO), "ISO 26262 Road Vehicles – functional safety" ISO, 2018.

[5] International Organization for Standardization (ISO), "ISO 21448 Road Vehicles – Safety of the Intended Functionality" ISO, 2022.

[6] International Organization for Standardization (ISO), "ISO PAS 8800:2023 Road Vehicles – Safety and artificial intelligence" ISO, 2023, CD draft.

[7] O. Willers, et al., "Safety Concerns and Mitigation Approaches Regarding the Use of Deep Learning in Safety-Critical Perception Tasks," Cornell University, 2020.

[8] C Miller, C Valasek, "Remote Exploitation of an Unaltered Passenger Vehicle", academia.edu, 2015.

[9] International Organization for Standardization (ISO), "ISO 21434 Road Vehicles – Cybersecurity engineering" ISO, 2021.

[10] International Organization for Standardization (ISO), "ISO 34502 Road Vehicles – Test scenarios for automated driving systems – Scenario based safety evaluation framework" ISO, 2022.

[11] UNECE, "Regulation No. 155 - Uniform provisions concerning the approval of vehicles with regards to cyber security and cyber security management system," 2021.

Image attributions:
top image: © Shutterstock id: 589267616
object detection: © HORIBA MIRA
car on chip: © Shutterstock id: 2187915949
networked vehicles: © Shutterstock id: 700959994

Dr Helen Monkhouse, HORIBA MIRA Ltd.

Dr Helen Monkhouse is a Chartered Engineer and system safety practitioner who has worked in the automotive domain for over 25 years. She is currently a Functional Safety Chief Engineer at HORIBA MIRA. Helen is a UK Expert to the British Standards AUE/32/1 Safety and Security of EE Components in Road Vehicles committee, and a co-author of MISRA's Guidelines for Automotive Safety Arguments.

The author retains copyright of this article.