Seminar: Safe Autonomous Transport – the Good, the Bad and the Ugly



The "Safe Autonomous Transport - the Good, the Bad and the Ugly" seminar was held at the Eurorstars Book Hotel in Munich, Germany on 28th November 2024 and hosted by Carmen Carlan and Mike Parsons. The event looked at progress in autonomous and highly automated transport and the problems encountered in implementing systems that are sufficiently safe.

This was the very first SCSC seminar held outside of the UK and was the inspiration of Carmen Carlan who felt that an event in Germany would better facilitate attendance from those, for various reasons, who were not normally able to travel to the UK. Carman was instrumental in organising the event, arranging the speakers and helping select a suitable venue. She also co-hosted with Mike Parsons, who, along with Alex King, Brian Jepson and myself, made the journey via train from the UK.

Destination: Munich!

Rather than fly separately, we decided to meet up at St Pancras in London and take the train as there was quite a bit of high-value AV equipment and we would be able to collaborate more closely travelling together.





The only problem was that the train journey entailed covering the best part of 1,000 miles for us all. Even if the trip had gone smoothly, we each had about 16 hours of travelling ahead of us! Unfortunately after

meeting up at St Pancras and a relatively smooth journey to Brussels, we arrived to find our Frankfurt train cancelled (we were scheduled to change there for the final leg to Munich). With not much information from staff on the ground and conflicting digital-based information on various European websites and apps, we decided to ignore the attendant's advice to travel to Aachan (just inside the German border) and wait for the next Frankfurt train. After lunch and feeling pleased with ourselves, we boarded the fast intercity ICE train at Liege Guillemins only to be told that the trip was going to be delayed 90 minutes and we would be better getting the regional train from Aachan! A further change at Cologne left us with a 5-hour journey to Munich and we eventually arrived at the hotel close to midnight, three hours late!





Markets, Soup and Museums!

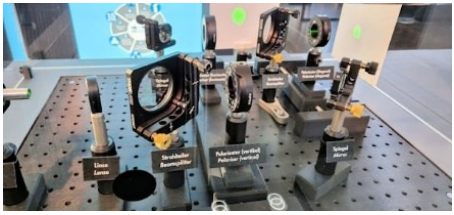
The team did have some time the next day to recover and see a few of the sights before setting up the room for the seminar the following day. Munich itself was very pleasant with a vibrant cosmopolitan feel enhanced by the Christmas market stalls that were marvellously lit in the evening. The stalls of course abounded with stollen, bockwurst, chocolate, cheese and Christmas stocking fillers, but there were also market stalls selling fruit and veg, giving the city a more local village feel, which was delightful. We eventually found a marvellous Suppenküche selling soups of course, but also chilli-con-carne and other dishes. Standing room only but very good and we rounded the meal off with some excellent coffees.





The rest of the day turned into a mini "Tech Trip" for Mike and myself. The Deutsches Museum sits next to the river Isar and is reputedly the world's largest museum of science and technology. We only had a few hours there, but it did not disappoint with many wonderful exhibits many being hands-on.

Particular interactive favourites for me were the quantum experiments such as the 'double-slit' and the 'information-eraser'. The museum also had a room dedicated to bridge building and architecture, which was fascinating and informative and the models used to illustrate building processes were mini works of art in their own right!





A model railway with a snowbound theme filled half of one room and there was a whole floor dedicated to aviation with more German-oriented craft such as the Messerschmitt and the Junkers with their distinct 'corrugated' bodywork.



We rounded the day off in the Ratskeller a very large and popular restaurant in the wonderful undercroft of the New Town Hall ('new' is perhaps a misnomer as it is over 150 years old now!) No meal is complete in Munich of course without a Bavarian Weissbier or two!

Trains, Planes and Automobiles



The seminar took place on the Thursday and was very well-attended with 32 in-person and 2 online. The Eurostars Book hotel was a very comfortable location with plenty of space for delegates and the refreshments and lunch were ample and of a high standard.



As the name suggests, the hotel has a literary theme with a marvellous Trompe L'oeil on the wall in the reception and many classic titles decorating the wall.

The six speakers presented a wide range of transport-related AI topics, touching on the domains of autonomous trains, aircraft and road vehicles:

- Mario Trapp, Safe Intelligence, Fraunhofer IKS, TUM - Assuring Safety in the Face of the Unpredictable
- Alex Haag, Futurail - How Safe is Safe Enough for Autonomous Trains? Analyzing Human Performance as a Reference System
- Levi Lúcio and Christoph Neuböck, Airbus - Lessons Learned from Enabling Model Based Systems Engineering for a Large Autonomous Aircraft Programme
- Torben Stolte, Volkswagen - ADMT's Approach Toward Arguing Safety for Automated Driving - An Introduction
- Simon Burton, University of York, UK - Safety Under Uncertainty: Automotive Standards for AI Safety and Research Perspectives
- Henrik Putzer, Cognitron - Developing & Assessing for Trustworthiness in AI



Mike Parsons introduced the event with an introduction to the SCSC and highlighted some of the perceived challenges of AI; not only in how we can assure systems, but how AI might need to be used to give us additional tools to help address the emerging challenges.

Mike then handed over to Carmen Carlan who introduced each of the speakers in turn.

Mario Trapp



Prof. Mario Trapp was our first speaker and started by describing the "Economic Safety Gap" where the excitement around the utility that AI can bring is exponentially driving technological innovation, but leaving an increasingly large gap between use and safe use. Safety engineers are also in a very small minority and so there is opposition around those that raise safety concerns.

Mario asked, what can we do to give safety a voice? He said we are a linear discipline now operating in an exponential world. Mario proposed that we need to lift safety to the next level as part of the next generation of safety. This he called 'Safety NxT' – a 5-point framework as summarised below:

Nos 1: To Change the Narrative

Traditionally, safety is in the nay-sayers domain and we are not good at storytelling. We need to tell our story in a more motivational and constructive way. Safety is a very small minority compared to all other stakeholders and so it's important to give one consistent voice for safety.

Nos 2: Design

Safety needs to be part of the design process from the beginning. A key point is the *value* that the system might bring and the value we want to create.

Nos 3: Speed

There is a desire to increase the speed of release of updates as technology is evolving exponentially. For example, Tesla has approximately 40 releases a year, so we can no longer take months to assess a release and so we need to be quicker than we are today.

Nos 4: Adaptability

Things that worked in the past may no longer work in the future, especially with AI. Typically in automotive, there is a 14-day cycle to recertify any change. This does however offer an opportunity to try things. For example, we can explore missing evidence gaps through gathering small pieces in the field and provide micro-evidence.

"We need to think about the value: the more capability we need, the harder it is to guarantee safety – if it is not safe there will be no value"

Nos 5: Adaptivity

What can systems do themselves to adapt? Mario's area of research is looking at how systems can be taught to adapt to an unpredictable context. Traditional safety assurance usually considers the worst case outcome for a hazard, but considering the current operational situation as part of a dynamic risk assessment at runtime will give freedom to optimise utility without violating safety.

Alex Haag



Alex was our second speaker. Alex works for the startup company Futurail, which aims to advance autonomous train technologies and AI products are seen as paving the way to full autonomy.

Alex asked why have autonomous trains at all and said the team were not targeting high volume trains like ICEs (less of a business case for relatively low number of vehicles), but more the thousands of regional trains where there is an acute driver shortage (5 drivers are required for each train due to working hour constraints). Replacing the drivers reduces costs not only in salaries, but also in the high cost of driver training and it can also increase line capability by optimising travelling speed.

Alex showed the high-level architecture of their Objective Detection System, which uses a combination of optical and LIDAR sensor data to detect objects. A 'structure gauge' is overlaid to show the safe space to be occupied by the train. He showed a video of the detection challenges and the need for filtering, for example, the train shouldn't brake if a pigeon flies in front of the train, and fog will introduce noise into the data. Alex said their systems are installed on a train in Belgium and have already capture 30,000km of very diverse data (eg. tunnels, narrow track, different seasons and climate conditions).

Vegetation detection is important as well as the rail itself and platforms can be challenging as, by design, they are very close to the structure gauge and so there is a need to distinguish things that are expected to be close and those that are not like a pedestrian.

Alex said their aim is to have a SIL 2 certified system by the end of 2026 in time for live train operations.

One question is how safe is safe enough? And what does SIL2 mean in this context and there are still discussions in these areas. The traditional methods to demonstrate safety (from EN50126) cannot be used:

- **Explicit risk estimation**: requires precise data that is not available (obstacle type, frequency in all environmental conditions)
- **Standard practices**: no Rail Object Detection Systems and AVs are all in US and China
- **Reference systems**: object detection is all human-based. There is accident data available but not avoided accidents

It is important therefore to estimate obstacle occurrence from unavoidable accidents, typically at night or in curves. Based on some assumptions, the team have calculated from accident data from trains running in South West France, that humans avoid 30% of accidents in these scenarios and so puts the target for an AI system into perspective, for example, maybe 90% detection rate for AI may be acceptable. Alex concluded by saying that this is still work in progress and more diverse data is being captured to help validate the system.

Levi Lucio and Christoph Neubock



Christoph presented remotely from Toulouse discussing MBSE in Airbus for a large autonomous aircraft programme running with a team of over 300 people. Christoph showed the V-model system engineering approach and showed the scope of the MBSE covering requirements definition through to design definition. The model covers all levels from the entire aircraft right down to equipment level.

Christoph said that because of the size of the team, there was a need for MBSE Model management to manage the combined efforts and contributions of 300 people.

This, for example, involves the assessment and measurement of the quality of the models. Even with everyone using MBSE, there is still a need to produce documents, including data and configuration management and there are security and IP rights to consider.

Levi then took over as lead presenter and said things have evolved since starting with MBSE and he shared some of the problems encountered, which occurred at many different levels:

Cultural Challenges: people were reluctant to discard the legacy processes that they were familiar with, there was missing knowledge/skills/ability for MBSE developments, missing understanding and still a heavy dependence on document-based ways of working.

Lessons learnt: value legacy (don't throw away as people are attached to it).

Scalability: there is a need to manage models at scale. Vendors' product features are not always as expected and it is hard to turnaround changes quickly. Levi said that they ended up writing software and tools themselves.

Lessons learnt: start with big bang but be flexible and make and introduce rules over time. Handle data exchanges and variants as first-class citizens.

Integration Challenges: There were issues with the integration of organisations, processes, methods and tools.

Lessons learnt: build the community on small and constant successes.

Software Tools: don't expect tooling to give you everything you need, but if you write it yourself, any software development has to be professional.

Lessons learnt: Mainly a matter of education on quality and insufficiency of external tools.

There are now, however, achievements of note gained through much pain!

Levi concluded by describing the road ahead. The first objective is to make sure the product can be architected (get the plane to work!) He thought they still need much more automation as there are still a lot of manual activities. More also needs to be integrated and many more lessons are expected to be learned!

Torben Stolte



ADMT's is VW's **A**utonomous **D**riving **M**aaS (mobility as a Service) and **T**aaS (Transport as a service). ADMT is their approach to arguing the safety of Level 4 Autonomous Vehicles, that is, AVs with no human drivers but operating in a limited area such as shuttle services and goods transport.

Torben said that we want AVs to be safe but we need a proper definition of "safe" as there is no universally agreed specification for the term. VW use the ISO26262 definition: "safety is the absence of unreasonable risk" but Torben asked how is unreasonable risk defined?

An engineering perspective might simply use a severity versus occurrences model but as Phil Koopman has said, 'safety is more than counting injured or dead people' and is much more nuanced. For example, there is biasing that may occur from particular road users such as cyclists. How also would targets be defined and how would these be modulated by the class of driver (eg. young, drunk, elderly etc.)

There is a need for safety argumentation but there are challenges:

- Proving, with statistical significance, events that are very rare would require many (billions) of miles to be driven, which is impracticable. Statistical proofs are not reasonably achievable prior to deployment.
- There is uncertainty about the safety level expected by the public, the perceived level of safety compared to the engineered level of safety, and perceptions are likely to evolve over time
- Regulatory demands are very coarse-grained at the moment

The approach is therefore to use safety indicators that collectively combine to form a claim that the system is free from unreasonable risk. Prior to operation, these will be prospective and include aspects such as good design and post-deployment they will be retrospective indicators using more statistical argumentation.

The Safety Assurance Case being developed is based on Goal Structuring Notation (GSN) and Torben provided a high-level GSN diagram of the argument but said structuring the argument is a multidimensional challenge: there are assurance domains over the system lifecycle but also cross-cutting topics that also need to be considered such as Safety Culture.

A key research question is how to express uncertainty in the Safety Assurance Case. For example how to address aspects such as qualitative versus quantitative approaches, argumentation and evidence uncertainty and the sheer diversity of topics.

Torben concluded with some observations:

- Safety argumentation must fit to the organisation
- Generalised argument patterns are hard to define due to diversity of topics
- Safety argumentation is an evolutionary process
- Top level safety claim formulation does not matter too much
- Terminology and consistent use of language is important

Simon Burton



Simon started his talk by stating that AI/ML models will eventually fail sometimes under certain conditions and asked a number of questions: how safe is safe enough? What measures should be taken to reduce risk? How do we argue an AI-based system is safe? Are there limits to what we can reasonably achieve and how do we interpret existing standard and regulations?

Simon said he was asked by ISO to lead the standardisation effort for AI so he first looked at existing standards. He found that ISO26262 focusses on malfunctioning behaviour but doesn't particularly address risks with non-malfunctioning but primarily erroneous behaviour. He also looked at ISO21448 (SOTIF), which requires the absence of unreasonable risk from functional insufficiencies and so might be a better standard for AI safety.

He then introduced ISO PAS 8800: "Road Vehicles – Safety and Artificial Intelligence". The intention was not to repeat requirements and guidance already held elsewhere so, for example, PAS 8800 does not cover functional safety arising from malfunction.

Simon provided an overview of the standard and said the draft was approved on 16th October 2024 and full release is scheduled for 13th December 2024.

He shared his thoughts on what was next for AI and safety and research directions and said one of the biggest open questions is:

Are we able to make definitive statements about the safety of AI?

He concluded that "it depends!" Simon defined uncertainty as "deviation from the unachievable ideal of completely deterministic knowledge of the relevant system" and showed a diagram that illustrated how there are a number of layers that compound uncertainty.

He said there was a relationship between the complexity of a system and our ability to make definitive statements about the safety of the system; the assurance argument therefore gets harder as the system gets more complex. He then showed how the standard could be applied by using the example of a construction sign detection system.

"Are we able to make definitive statements about the safety of AI? ... it depends!"

Simon then presented his view on the limitation of safety assurance providing examples of applications as they sit on a graph of complexity of environment/task/data/model versus assurance uncertainty with some "Frontier AI" for safety-critical applications now being suggested such as using LLMs (self-trained on the Internet) for AV driving tasks.

He then listed the fundamental challenges that need to be overcome:

Can AI still help in managing the complexity of systems? Should we really use LLM which are simply "guessing the next word" in safety systems or can they be used in a more hybrid fashion? Simon concluded with a summary of the foundations of an AI safety argument (found in the standard but still evolving).

Henrik Putzer

The final presentation was from Henrik covering Developing and Assuring for Trustworthiness in AI. He structured his talk around the elements in the seminar's title: The Good, The Bad and The Ugly.



The Good

AI is quick to market and there is commercial imperative to include it in future products from company's like Bosch. He said there is a lot of AI going on and it will undoubtedly be part of our systems in some form in the future.

Henrik asked what is so special about AI and looked at the history of technology usage from mechanics, electronic and software and now arriving at AI/ML/DL. AI can make decisions about larger unseen datasets based on only partial samples of training data. However, it also introduces uncertainty related failures and Henrik gave some of example of 'AI Fails'.

"We shouldn't be afraid of AI but afraid of those that don't use it properly"

The Bad

There is Mysticism about AI, which has been seen in research – changing the order of training data sets or quantifying a neural net model can suddenly make everything work. We need to change Mysticism using structured processes.

Henrik then looked at the timeline of functional safety and asked whether the old risk-based approaches can be used in AI? Henrik thought so, but with some tweaks, in particular there is a need to consider fault models in the AI.

In summary AI still sits within the system architecture and covered by standards like ISO 26262. A risk-based approach can still be adopted so it's important to understand the AI failures through a fault model and there needs to be a structured approach, so a V-model for AI including specific methods to be applied.

Henrik then discussed a standard that introduces a structured approach to developing Deep Neural Network (DNN) elements (VDE-AR-E 2842-61 part 5). The approach is based on a V-model. The left-hand side has the Design phases where the training dataset is built. After hyperparameter optimisation there is then a series of verification activities on the right-hand side with training, design, functional and semantic verification.

The Ugly

Henrik said "The Ugly" side of AI from an engineer's perspective is when the assessors come along! However, by following the standard the situation should not be so bad as embedded in the V-model are the general process areas that an assessor will be looking for.

In summary Henrik said that he believes there will be (and is already) beneficial (and safe) AI but we have work to do and there needs to be system (of systems) engineering to support it. We shouldn't be afraid of AI but afraid of those that don't use it properly.

Mike Parsons – Wrap Up

Mike wrapped up the seminar with a general discussion session, which allowed the attendees to discuss their thoughts on specific AI-related questions such as:

> *Question: Do we have the whole ecosystem we need for AI covering all aspects such as development, evolution, regulation, maintenance and documentation?*

Mike then opened the floor for any further discussion topics.

Conclusions

Despite some of the epic train journeys required to bring the event to mainland Europe, the seminar was hugely successful with fantastic speakers in a great venue. The event being so well-attended led to some great discussions during the breaks and it was a fantastic opportunity to share ideas and current best practice in the challenging area of AI in transport systems.



Feedback from delegates on the event have been very positive and so we hope this will be the first of many more seminars to be held on mainland Europe.

Report by Paul Hampton SCSC Newsletter Editor

Images: © Paul Hampton and Mike Parsons

