

The Safety-Critical Systems Club Newsletter

# Safety Systems

Vol 32 No. 2 – May 2024

## THE GRENFELL LEGACY
Are we making buildings safer?

## BURDEN OF PROOF
Learning from the Policing domain

## WHAT LIES BENEATH?
Is something lurking in your code?

## BOUNDARY CONDITIONS
The expanding scope of AI Safety

Safety Critical Systems Club Supporting Development of Safer Systems
The Systems Safety Community and Club
SCSC

SCSC

thescsc.org




SCSC Publication Number: SCSC-194

While the authors and the publishers have used reasonable endeavours to ensure that the information and guidance given in this work is correct, all parties must rely on their own skill and judgement when making use of this work and obtain professional or specialist advice before taking, or refraining from, any action on the basis of the content of this work. Neither the authors nor the publishers make any representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability or availability with respect to such information and guidance for any purpose, and they will not be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever (including as a result of negligence) arising out of, or in connection with, the use of this work. The views and opinions expressed in this publication are those of the authors and do not necessarily reflect those of their employers, the SCSC or other organisations.

© SCSC CIC 2024 except where explicitly stated for specific articles, reports or images.

Front cover and editorial: AI generated images
In Brief images on page 4 may be subject to copyright – see scsc.uk/f138 for details.

# Contents


## WELCOME

**Editorial** 3

Opening words from the SCSC Newsletter Editor.

**In Brief** 4

Recent system safety news items from around the world.

## FEATURES

**Building Safety Post Grenfell** 5

Laura Mills discusses the progress made since the Grenfell Tower fire in 2017

**Evidence Management – A Policing Perspective?** 13

Bill Blackburn discusses what we can learn from the Policing domain.

**Is there something hidden in the code?** 21

Gavin Wilsher discusses how to mitigate undesirable code appearing in systems.

**AI Safety** 27

Ibrahim Habli and John McDermid discuss the expanding scope of AI safety.

## PERSPECTIVES

**SCSC Statement on the Post Office Horizon** 2

An SCSC Steering Group statement.

**Letters to the Editor** 33

A new feature covering letters sent to the SCSC Newsletter Editor.

## REPORTS

**SSS'24 Event Report** 35

Highlights from this year's Safety-Critical Systems Symposium.

**Poster Sessions** 51

Copies of the posters presented at SSS'24.

**Recent Safety Publications** 70

The latest publications relating to safety.

**60 Secs with ... Carmen Carlan** 71

Carmen answers some quick-fire questions on system safety and life!

## GROUPS

**Working Groups** 74

Details of the SCSC Working Groups.

**SCSC Steering Group** 80

Who's who in the Steering Group.

## EVENTS

**Calendar** 82

**Events Diary** 83

# SCSC Statement on the Post Office Horizon

The Post Office Horizon IT case has been subject to an inquiry, due to conclude this year, which is examining the Post Office's prosecution of over 700 sub-postmasters for theft, false accounting and related charges because of technical faults in the Horizon IT system. These accusations persisted for 16 years – from 1999 to 2015 and has been described as one of the greatest miscarriages of justice in British history. Given the scale of the issues, the role of IT and the clear safety implications that have arisen, the SCSC Steering Group would like to make the following statement:

- The Post Office Horizon system would not normally be regarded as a safety system, yet it is a computer-based system that has indirectly led to widespread harm.
- The ongoing public inquiry has raised important legal, ethical and technical concerns. Problems highlighted include: sub-postmasters could not see what was going on (for example, figures changed remotely without sub-postmaster knowledge), poor quality coding and lack of both audit and fault logging.
- The SCSC fully supports the public inquiry and other investigations, and we agree with professional computer bodies (e.g. the British Computer Society) that there should be a review of how computer-based system evidence is treated by the courts.
- We will look to adopt relevant recommendations and encourage our members to do the same once the inquiry and investigations conclude.
- We propose that organisations relying on computer-based system evidence in court should, where challenged, be required to justify that the system, including aspects such as hardware, software, data and service delivery, is reliable. Furthermore the evidence should be shown to be trustworthy. The justification should also show appropriate confidence in use, including in the way that reported problems are managed. Courts should not accept evidence relating to the computer-based system without this justification.
- Where there are disputes involving computer-based systems there must be fair treatment; i.e. where relevant, there must be access for both sides to technical experts who in turn must be given access to appropriate software and data.
- Post Office Horizon is an example of how systems, organisations, agreements, people and processes came together within a delivered service to result in indirect but severe harm. We suggest that the SCSC Service Assurance Guidance could be useful in such situations to reduce risks.
- We will extend the remit of the SCSC to cover any computer-based systems and services which could cause harm. This will include harm in the wider context of the system including all stakeholders, the environment and consequential harms, not just harm caused directly by the system or service itself.

Note: Here "computer-based system" includes aspects such as hardware, software, data and service delivery.

# Editorial

This edition of the newsletter, being the first to follow the SCSC's symposium in February, gives us time to reflect on that landmark event and it is my great pleasure to provide highlights of proceedings in my report (page 35) along with copies of the posters displayed (page 51) during the interactive session on the second day. You can access a digital version of each poster by clicking on each image in the newsletter.

The event was very well-attended and it was great to see so many diverse and interesting talks. One striking talk from Prof. Harold Thimbleby highlighted the issues surrounding the Post Office Horizon scandal, arguably the worst miscarriage of justice in UK history. As with other organisations, the SCSC Steering Group has been moved to produce a statement on this incident that outlines what action the group believes organisations (including the SCSC itself) should take to ensure such events are not repeated in the future. You can see the full text of the statement opposite.

Artificial Intelligence (AI) was an important theme for SSS'24 with some excellent talks and a dedicated AI Panel. If you missed this then I would encourage you to catch up with the video recording on the sessions to see the responses to various burning questions around AI such as: "How do we make AI safe?" and "What might limit the arguably unbridled expansion of AI systems?" SSS'24 also saw the formal launch of the SCSC's Safe AI Working Group (SAIWG). The group aims to share knowledge and experience on various topics related to AI and safety systems, such as coordination of safety with other disciplines, evaluation of risk, and mapping of terminology and language. Please see the Working Group descriptions (starting on page 74) for more details on the group and how you can get involved.

We have also already embarked on a series of seminars relating to AI, and I hope to include a report of the first one: "Frameworks for Safe AI Systems" in the next edition of the newsletter. There is still time to book the remaining events with "Developing Safe AI Systems" being held in London on 27th June 2024 and "Deployment, Operations and Maintenance of Safe AI Systems" being held also in London on 26th September 2024. We are also very excited to announce our intentions to hold an international seminar on the safety of autonomous vehicles in Munich, Germany on 28th November 2024, so please watch out for further announcements on that event.

There have been a few changes in the Steering Group so I have pleasure in welcoming Yvonne Oakshott and Carmen Carlan both as new members to the group. You can find out a little bit more about Carmen in my 60 second interview with her on page 71. We also welcome Khadijah Khatun as new lead for the Safety Futures Initiative (SFI) to carry on the good work already done by Zoe Garstang and Nikita Johnson.

In this edition we have four great and diverse articles covering Building Safety, Policing, Railways and Artificial Intelligence (AI). I am also introducing a new 'Letters to the Editor' section (see page 33) where I will publish letters I receive on any matter relating to systems safety. Our first is from Dewi Daniels and I would invite you all to write in with any views and opinions for this section to be included in future editions of the newsletter.

**Paul Hampton**
**SCSC Newsletter Editor**
**paul.hampton@scsc.uk**



# In Brief



## FAA audit of Boeing's 737 production found mechanics using hotel card and dish soap as makeshift tools




The Federal Aviation Administration found dozens of issues throughout Boeing's 737 MAX jet production process, including mechanics at one of its key suppliers using a hotel key card and dish soap as makeshift tools to test compliance, according to a report.

The FAA discovered "unacceptable" quality control issues during an audit of Boeing and supplier Spirit AeroSystems that was launched after a door plug flew off a 737 MAX 9 at 16,000 feet on 5th Jan.
*nypost.com*

## Top Execs at Toyota's Daihatsu Bow Out Over Safety-Test Scandal




Top executives at Toyota Motor's compact-car unit Daihatsu will step down, weeks after the carmaker suspended all shipments due to a safety scandal investigation. In April last year, Daihatsu said it had rigged side-collision safety tests carried out for 88,000 small cars, most of which were sold as Toyotas.

By December, however, investigators had found that the scope of the scandal was far greater and went back much further, affecting Toyota's Mazda and Subaru brands as well.
*asiafinancial.com*

## Six presumed dead after cargo ship crash levels Baltimore bridge



A major Baltimore bridge collapsed like a house of cards early Tuesday after it was struck by a container ship, sending six people to their deaths in the dark waters below, and closing one of the country's busiest ports.

The tragic chain of events began early Tuesday when the cargo ship Dali notified authorities that it had lost power and issued a mayday moments before the 984-foot vessel slammed into a bridge support at a speed of 8 knots.
*nbcnews.com*

## Mesh and sodium valproate scandal victims need payouts soon, report says




Families of children left disabled by an epilepsy drug and women injured by pelvic mesh implants should be given urgent financial help, England's patient safety commissioner has said.

Dr Henrietta Hughes has called on the government to act quickly to help victims of the two health scandals.
*bbc.co.uk*

# Building Safety Post Grenfell

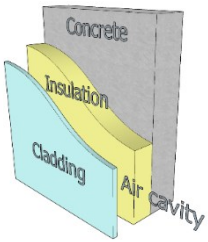


**In 2017 the world witnessed the horrific tragedy of the Grenfell Tower fire that claimed the lives of 72 people making it one of the deadliest UK residential fires. This resulted in a number of public enquiries and a raft of legislative changes aimed at improving safety of high-rise buildings both in new and existing developments. Laura Mills describes the legislative changes and assesses what progress has been made in ensuring fires of this magnitude never occur again.**

## Introduction

The Grenfell Fire Tower in 2017 was a huge tragedy where 72 people lost their lives. The investigations that followed included the independent review of building regulations and fire safety (The Hackitt Review); this was a deep dive into the construction industry and its regulation and resulted in Dame Judith Hackitt highlighting Building Safety as a global concern. The second investigation was the Grenfell Tower Inquiry, led by Sir Martin Moore-Bick, to examine the fire and also the circumstances which led up to the fire. The combined recommendations from the inquires led to reform in legislation within the built environment and the regulatory regime.

## Cladding




It is widely reported upon that when the fire escaped from flat 16 and entered the cladding, it spread rapidly up the east face of the tower, then around the top of the building in both directions and down the sides, enveloping the entire building in under three hours. Aluminium Composite Material (ACM) rainscreens with polyethylene cores acted as a source of fuel and one of the main reasons for the horizontal and downward spread of the fire was the melting and dripping of burning polyethylene from the spandrel and column panels that ignited fires lower down the building [2].

Many other existing buildings have similar cladding and so they are exposed to similar risks in the event of fire. The response to this issue was a Cladding Safety Scheme (CSS) funded by both the £5.1 billion allocated by the government to fix the most dangerous buildings and through revenue from the Building Safety Levy on new development. The scheme available to all medium-rise buildings between 11 and 18 metres across England and high-rise buildings over 18 metres (outside of London) where fire safety professionals have recommended that works must take place. The scheme is also available to the social housing sector with the aim of protecting leaseholders from costs where the responsible developer cannot be made to pay [10].

## Issues relating to Fire & Building Safety (Grenfell Inquiry Phase 1)

The Grenfell Tower Inquiry Phase 1 report highlighted the lack of on-site information available to the London Fire Brigade on attendance. While some of the senior officers attending had an awareness of the danger of cladding fire, others were not trained to recognise the dangers of these systems and how fire development and fire behaviour could occur. If the building design had been more complex, the lack of information available could have had further impacts [2].

One of the recommendations from the inquiry centred around the availability of plans of high-rise residential buildings to local fire and rescue services and the provision of premises information boxes in high-rise residential buildings. In response to these recommendations, The Fire Safety (England) Regulations 2022 [5], implemented a change in the law for high-rise residential buildings (a multi-occupied residential building at least 18 metres in height or 7 or more storeys) – responsible persons must:

- Share electronically with their local Fire and Rescue Service (FRS) information about the building's external wall system (Regulation 5) and provide the FRS with electronic copies of floor plans and building plans for the building (Regulation 6) [5]
- In addition, hard copies of the building's floor plans, a single page orientation plan of the building, and the name and UK contact details of the responsible person must be held in a secure information box (Regulation 4), which is accessible by firefighters [5]

This knowledge and information will aid FRS teams in planning for emergency response situations.

- Additionally, Section 1 of the Fire Safety Act 2021 also now clarifies that where a building contains two or more sets of domestic premises, the responsible person must take account of structure and external walls in a fire risk assessment. The Fire Safety Act amends article 6 of the Regulatory Reform (Fire Safety) Order Act 2005 [4].

When fire-fighters attended the fire at Grenfell, they were unable to operate the mechanism that should have allowed them to take control of the lifts. They were therefore unable to make use of the lifts, which are often key to carrying out firefighting and search & rescue operations successfully. Occupants of the tower were also unable to use the lifts to escape, in some cases, with fatal consequences [2]. Grenfell tower was 24 storeys high, equal to 67.3m or 221ft [6]. The inquiry recommendation for regular inspection and testing of lifts designed for use by firefighters [2], has been enforced via The Fire Safety (England) Regulations 2022 [5] under regulation 7. Regulation 7 requires monthly inspections of not only lifts, but all key fire-fighting equipment provided in the building to fight fires and reporting any defects to local fire and rescue service electronically within 24-hours. These regulations will help prevent similar situations as the Fire and Rescue Service, in the majority of cases, will know before arrival whether a lift intended for their use in an emergency is operational.

Inquiry recommendations also included the provision of fire safety information to residents of high-rise residential buildings and the marking of floor levels in lobbies and staircase landings [2]. In the Grenfell building, the staircase landings were not marked, firefighters were unable to identify floors clearly when carrying out firefighting or search and rescue operations within the building [5]. The Fire Safety (England) Regulations 2022 sets out the requirement under regulation 8 for floor numbers to be clearly marked on each landing within the stairways, and in a prominent position of all lobbies, which should be visible in normal conditions and in low lighting/smoky conditions [5] to avoid this reoccurring. This recommendation had previously been raised within the Coroner's Rule 43 letter following the fatal fire where two fire fighters tragically died at Shirley Towers in Southampton in 2010 [7].

Evidence from the Inquiry [2] indicated many occupants of Grenfell Tower were unable to read or understand the fire safety instructions placed in the lobbies throughout the building. The Fire Safety (England) Regulations 2022 now sets out the requirement under regulation 9 for the provision of fire safety instructions, including evacuation instructions, in a form that the occupants of the building can reasonably be expected to understand. This relates to all multi-occupied residential buildings in England with two or more sets of domestic premises over 11m or four stories (and which have common parts) [5] to help save lives.

The Inquiry highlighted a need for the inspection of fire doors and self-closing devices due to evidence obtained from Grenfell that defective fire doors allowed smoke and toxic gases to spread through the building more quickly [2].

- Regulation 10 of The Fire Safety (England) Regulations 2022 again relates to all multi-occupied residential buildings over 11m and requires routine checks and the provision of information to resident relating to fire doors
- In addition to this, Section 1 of the Fire Safety Act 2021 also now clarifies that: where a building contains two or more sets of domestic premises, the responsible person must take account of flat entrance doors in a fire risk assessment. The Fire Safety Act amends article 6 of the Regulatory Reform (Fire Safety) Order Act 2005 in this regard [4].

## The Hackitt Review




The interim report identified that the current system of building regulations and fire safety was not fit for purpose and that a culture change is required to support the delivery of buildings that are safe.

The Hackitt review has had an astounding impact on building regulations and fire safety from the focus of its initial parameters, this relates to both the built environment (the physical structures that humans inhabit) and buildings currently in occupation [1].

## The Built Environment

Recommendations set out to improve the focus on building safety during their design, construction and refurbishment whilst also tackling poor procurement practices & competence, all under a new framework.

Building (Higher-Risk Buildings Procedures) (England) Regulations 2023

The new building control process for higher-risk buildings in England came into force on 1st October 2023 via the Building (Higher-Risk Buildings Procedures) (England) Regulations 2023. From this date, the building control authority for higher-risk buildings is the Building Safety Regulator and new building work in scope can no longer be overseen just by local authority or private sector building control. This includes:

- Construction of new higher-risk building (HRB)
- Building work to existing HRBs
- Building works that change existing buildings to become HRBs (including material change of use)
- Building works that will declassify a building from being an HRB [8]

The Building Regulations etc. (Amendment) (England) Regulations 2023

To raise standards across the built environment, these regulations apply to all building work, with legal responsibilities placed on those who commission building work, participate in the design and construction process and carry out the building control function, to ensure compliance with building regulations [8].

The Building (Approved Inspectors etc. and Review of Decisions) (England) Regulations 2023

Make amendments to the Building (Approved Inspectors etc.) Regulations 2010 to begin to implement the changes introduced by the Building Safety Act 2022 and support the new building control regime. Changes by focus on competence and accountability by establishing a regulated building control profession [8].

## Occupied Buildings

The Hackitt report noted the need for duty holders with specific responsibilities, regulation, improvements relating to building information and that residents need to be given a voice, all under new regulation that hold powers of enforcement.

**"The Hackitt report noted ... that residents need to be given a voice, all under new regulation that hold powers of enforcement."**

The Building Safety Act 2022

Granted Royal Assent on 28th April 2022.

- Part 2 established the role of Building Safety Regulator and Part 4 includes information relating to its enforcement powers and relevant appeal processes
- Part 3 initialises changes to building control authorities and building regulations, including building control [9]

Part 4 comprehensively addresses all improvements for in-occupation higher-risk buildings (defined as those which meet 7 stories or 18 metres), and defines the role of duty holders with specific responsibilities that include:

- Registration of all existing buildings with the Building Safety Regulator in 2023, and the long-term requirement for obtaining building assessment certification from the regulator
- Prevention of building safety risks, defined as 'spread of fire and/or structural failure' via assessment, management, notification, and the creation of Building Safety Cases [9].

Safety Cases

Safety cases have been used in Health & Safety since the 1970s and the Building Safety Regulator has adopted these as part of Building Safety. Safety Case Reports should now demonstrate a building is safe for occupation and demonstrate that:

**"Safety Case Reports should now demonstrate a building is safe for occupation"**

- Building safety risks in relation to the building have been assessed
- All reasonable steps have been taken to prevent the occurrence, and control the impact of, any major incident resulting from building safety risks
- There is compliance with the prescribed principles introduced via secondary legislation
- All obligations in relation to resident engagement and participation, complaints handling, information provision and the role of residents in helping keep the building safe have been met
- A Resident's Engagement Strategy has been compiled and is maintained

Building Assurance Certificates will then be issued where the above can be evidenced [9].

**"a 'Golden Thread of Information' – this is the information about a building that enables it to be kept safe and is accurate, easily understandable, up-to-date and readily accessible"**

Golden Thread

Part 4 also includes the duty holder' responsibilities relating to documents and information:

- Ensuring information is suitably stored, provided to the regulator and residents where required, and where required, passed over to future accountable persons and a mandatory incident reporting mechanism for HRBs.

An additional parameter of the Hackitt Review was obligating the creation of a digital record, the information and documents required for all new and occupied HRBs to help create a 'Golden Thread of Information' – this is the information about a building that enables it to be kept safe and is accurate, easily understandable, up-to-date and readily accessible. Information and documents are required to be held digitally and established at the start of construction. The Golden Thread should be updated and maintained throughout a building's lifecycle.

Residents' duties and engagement

Residents' duties and engagement are also covered in Part 4:

- Resident engagement strategies are required, to provide greater transparency of information on building safety and residents must be advised of how to obtain further information
- New routes for complaints should be provided by the accountable person(s) and also the regulator to ensure concerns can be voiced
- Residents should also be made aware of duties placed on them, contravention notices and access to premises that are applicable to their responsibilities [9]

## What's next – Grenfell Inquiry - Phase 2

In Phase 2 the Inquiry (recently postponed until after the 7th anniversary of the tragedy) will seek to answer various questions set out in the list of issues, some of which are additional matters of particular concern relating to fire & building safety [2].

- The decisions relating to the design of the refurbishment and the choice of materials
- The regime for testing and certifying the reaction to fire of materials intended for use in construction
- The design and choice of materials
- The performance of fire doors
- The warnings of potential fire hazards given by the local community
- The width of the stairs
- The supply of gas
- The supply of electricity and the history of electrical surges

## Conclusions

The Grenfell Tower fire was one of the deadliest UK residential fires and rightly resulted in a number of public enquiries and a raft of legislative changes aimed at improving safety of high-rise building both in new and existing developments.




Whilst there have been many changes in the last seven years since the tragedy, it seems there is a long way to go. Inside Housing reported in March 2024 that 'of 3,839 buildings above 11 metres being monitored by the government due to the need for cladding remediation, 2,286 have not even started works yet' [11]. As we approach the seventh anniversary, we are only just starting to see the initial requests for safety cases and requirements relating to the registration of building inspectors becoming enforceable; this will be a point of interest to follow.

An enforceable obligation relating to signage was recommended following the fatal fire at Shirley Towers prior to Grenfell, this was subsequently re-raised within the Grenfell Inquiry and highlights situations which arose at Grenfell could have been avoided. Reports on Action to Prevent Future Deaths (under Schedule 5 of the Coroners and Justice Act 2009 and Regulations 28 and 29 of the Coroners (Investigations) Regulations 2013) [7], previously known as a Coroner's Rule 43 letter are compiled by the Coroner following fatal fires, where an investigation gives rise to concern that future deaths will occur. Whilst the corner raises issues and identifies specific areas of concern, with the recommendation that action should be taken, this does not include what that action should be, this remedy is for the person or organisation to whom the report is directed [12]. I wonder if there will now be changes in the way Action to Prevent Further Future Deaths will be handled [7].

The Grenfell Tower had a profound impact on its residents and indeed all of us have felt the impact on some level. It is important residents can feel safe and to know they have a voice in maintaining that safety, and that voice is being heard. Residents should therefore be at the heart of all of these changes both as the primary focus for safety and in ensuring transparency and collaboration on proposed changes.

# References


[1] Hackitt, J. and Freng, D. (2018). Building a Safer Future Cm 9607 Independent Review of Building Regulations and Fire Safety: Final Report. [online] Available at: https://assets.publishing.service.gov.uk/media/5afc50c840f0b622e4844ab4/Building_a_Safer_Future_-_web.pdf

[2] Moore-Bick, M. (2017). GRENFELL TOWER INQUIRY: PHASE 1 REPORT OVERVIEW REPORT of the PUBLIC INQUIRY into the FIRE at GRENFELL TOWER Chairman: The Rt Hon Sir Martin Moore-Bick. [online] Available at: https://assets.grenfelltowerinquiry.org.uk/GTI_-_Phase_1_report_Executive_Summary.pdf, accessed April 2024

[3] Sheldon, C. (n.d.). To: (England only) Local Government Chief Executives and Heads of Building Control Approved Inspectors The Chief Fire Officer Fire and Rescue Authorities. [online] Available at: https://assets.publishing.service.gov.uk/media/65bb9459cc6fd6000d5dbe4f/Building_Circular_Letter_-_Changes_to_the_building_control_process_update_January_2024.pdf, accessed April 2024

[4] Home Office (2022). Fire Safety Act 2021. [online] GOV.UK. Available at: https://www.gov.uk/government/publications/fire-safety-act-2021, accessed April 2024

[5] GOV.UK. (2022). Fire Safety (England) Regulations 2022. [online] Available at: https://www.gov.uk/government/publications/fire-safety-england-regulations-2022, accessed April 2024

[6] www.skyscrapercenter.com. (n.d.). Grenfell Tower - The Skyscraper Center. [online] Available at: https://www.skyscrapercenter.com/building/grenfell-tower/30139, accessed April 2024

[7] GUIDANCE No.5 REPORTS TO PREVENT FUTURE DEATHS 1. (n.d.). Available at: https://www.judiciary.uk/wp-content/uploads/2013/09/guidance-no-5-reports-to-prevent-future-deaths.pdf, accessed April 2024

[8] GOV.UK. (n.d.). Criteria for being a higher-risk building during the occupation phase of the new higher-risk regime. [online] Available at: https://www.gov.uk/guidance/criteria-for-being-a-higher-risk-building-during-the-occupation-phase-of-the-new-higher-risk-regime, accessed April 2024

[9] gov.uk (2022). Building Safety Act 2022. [online] Legislation.gov.uk. Available at: https://www.legislation.gov.uk/ukpga/2022/30/contents/enacted, accessed April 2024

[10] GOV.UK. (n.d.). Government launches biggest cladding removal scheme. [online] Available at: https://www.gov.uk/government/news/government-launches-biggest-cladding-removal-scheme, accessed April 2024

[11] Inside Housing. (n.d.). Cladding remediation: why is progress so slow? [online] Available at: https://www.insidehousing.co.uk/insight/cladding-remediation-why-is-progress-so-slow-85520, accessed May 2024

[12] Courts and Tribunals Judiciary. (2020). Revised Chief Coroner's Guidance No.5 Reports to Prevent Future Deaths[i]. [online] Available at: https://www.judiciary.uk/guidance-and-resources/revised-chief-coroners-guidance-no-5-reports-to-prevent-future-deathsi, accessed May 2024



Image attribution
leading image: https://commons.wikimedia.org/wiki/File:Grenfell_Tower_fire_%28wider_view%29_%28cropped%29.jpg, author: Phoenix7777, licensed under CC Attribution-Share Alike 4.0 International.
cladding: https://upload.wikimedia.org/wikipedia/commons/b/bf/Grenfell_Tower_cladding.png, licensed under CC 4.0
Hackitt Inquiry front cover: https://assets.publishing.service.gov.uk/media/5afc50c840f0b622e4844ab4/Building_a_Safer_Future_-_web.pdf, public domain.
United for Grenfell: Id 250573766 | Grenfell © Jessica Girvan | Dreamstime.com



**Laura Mills BSc AIFireE MCIOB**

Laura is an experienced Surveyor & Project Manager currently specialising in Building Safety. Skilled in Surveying, Planning, Project Management, Operations Management and Contractor Management.



The author retains copyright of this article

# Evidence Management – A Policing Perspective



**Bill Blackburn provides an understanding of the operational practices in policing where the establishment of reliable evidence is critical for prosecutions. Bill draws parallels with the evidential requirements to support a safety case for safety-critical systems and asks what we can learn from policing practices, such as how we can better position ourselves should a safety incident result in a more formal investigation, potentially leading to court.**

The Safety-Critical Systems Club (SCSC) ran an event – What Makes Good Evidence for a Safety Case? – on 30th November 2023, held at the Wellcome Collection in London.

Speakers and discussions provided the participants with an opportunity to reflect on the best practices and pitfalls across a number of fascinating exemplars. With my background in high integrity systems and my experience gained through working within the police and justice sectors, I could see many parallels that are worth exploring.

This article emerged from that event and sets out in broad terms how evidence management operates in the UK and most other jurisdictions. It aims to provide an understanding of operational practice in policing and where there are opportunities to align safety engineering methods. This cross-fertilisation is hopefully useful in its own right but also to put ourselves in a better position should a safety incident result in a more formal investigation, potentially leading to court.

## Context for handling evidence in policing

Police gather evidence as a result of an incident occurring and generally a victim making an allegation of a crime. A crime is something that is defined for a particular jurisdiction (e.g. England and Wales) to be an offence subject to investigation, prosecution and dealt within the context of one or more court hearings.

An offence is defined in legislation which is translated into offence wording and aspects such as mode of trial and sentencing provisions. Let's take criminal damage as an example. The Criminal Damage Act (CDA) 1971 has sections setting out a range of situations in which criminal damage might be done, along with various aggravating factors.

The legislation is subject to further codification to allow police, prosecution and courts to put this into practice. For the purpose of investigation and subsequent case building for decisions on whether the crime report can be progressed, the key element is an understanding of the "points to prove". Once that is clear, then the investigation focus is on gathering evidence in relation to these points.

Expanding on CDA 1971, the investigation and prosecution need to be able to prove that person A damaged a specific item of property (smashed a window of a garden shed at 38 Acacia Avenue) belonging to person B by some means (by throwing a brick from nearby) at a point in time (around 3:30 Wednesday afternoon) with an intent to cause damage or by being reckless. To make it easy from an information management viewpoint this can be considered as made up of Persons, Objects, Locations & Events (POLE). In our example:

- **P**ersons (A & B)
- **O**bjects (Garden Shed & Brick)
- **L**ocations (38 Acacia Avenue)
- **E**vents (Between 3pm and 4pm Wednesday afternoon)

All very straightforward then! But keep in mind UK legislation stretches back many years, with common law, bye-laws, subject to amendments and repeals, case precedence and other complicating factors. New legislation is enacted and the implications on prior legislation needs to be taken into account.

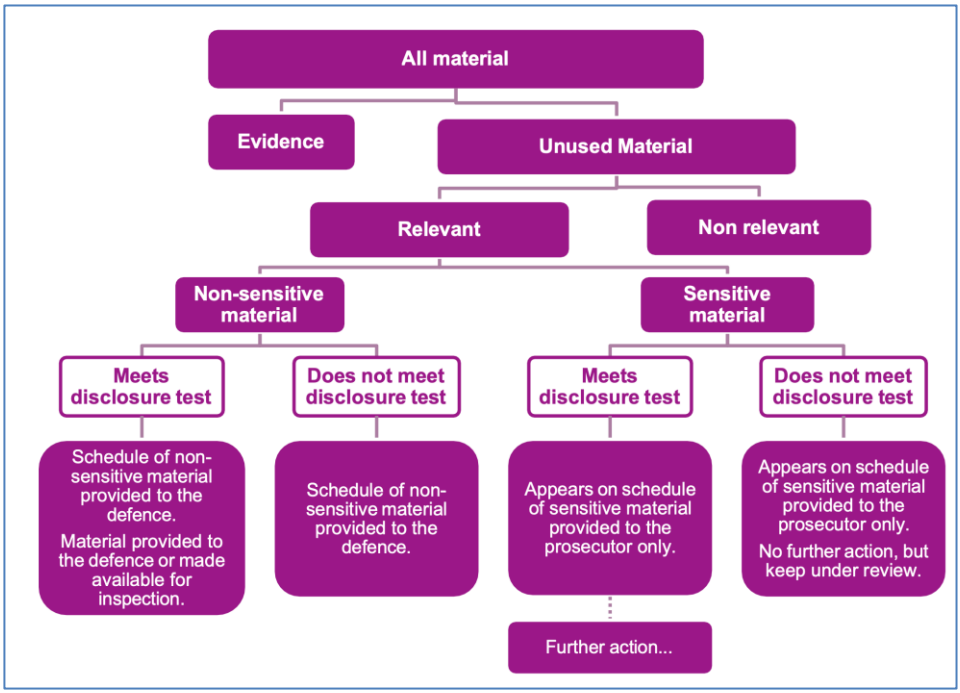
## What is Evidence in Criminal Justice?

During the course of an investigation numerous lines of enquiry may be followed, statements taken and material gathered. However the concept of evidence relates to the offence that is being investigated. Further legislation defines clear rules for how investigations are done and how evidence is managed, such as the Police and Crime Evidence Act 1984 (PACE) and Criminal Procedure and Investigations Act 1996 (CPIA). These are operationalised through training, guidelines and codes of practice.

**"If challenged in court, material not appropriately managed … can result in a case being dismissed."**

The Attorney General's Guidelines on Disclosure (AGO Guidelines) [1] were released in 2020 and updated regularly since to further clarify how evidence is managed to align with CPIA. The driver for this is that if challenged in court, material not appropriately managed and the right decisions not made for disclosure to the defence, can result in a case being dismissed, despite there being sufficient evidence. A suspect/defendant might simply be able to walk free and justice is not served.

The figure below is taken from the AGO Guidelines and provides a clear taxonomy for what is, and what is not, regarded as evidence but has been acquired during an investigation. It is worth bearing in mind that an initial view of evidence and unused material might change as new information comes to light. Also unused material (i.e. the prosecution is not relying on this) might become evidence that is used by defence to counter the prosecution. Dealing with evidence is in practice "Material Handling", at which point, the complexities multiple and the volume of material expands.

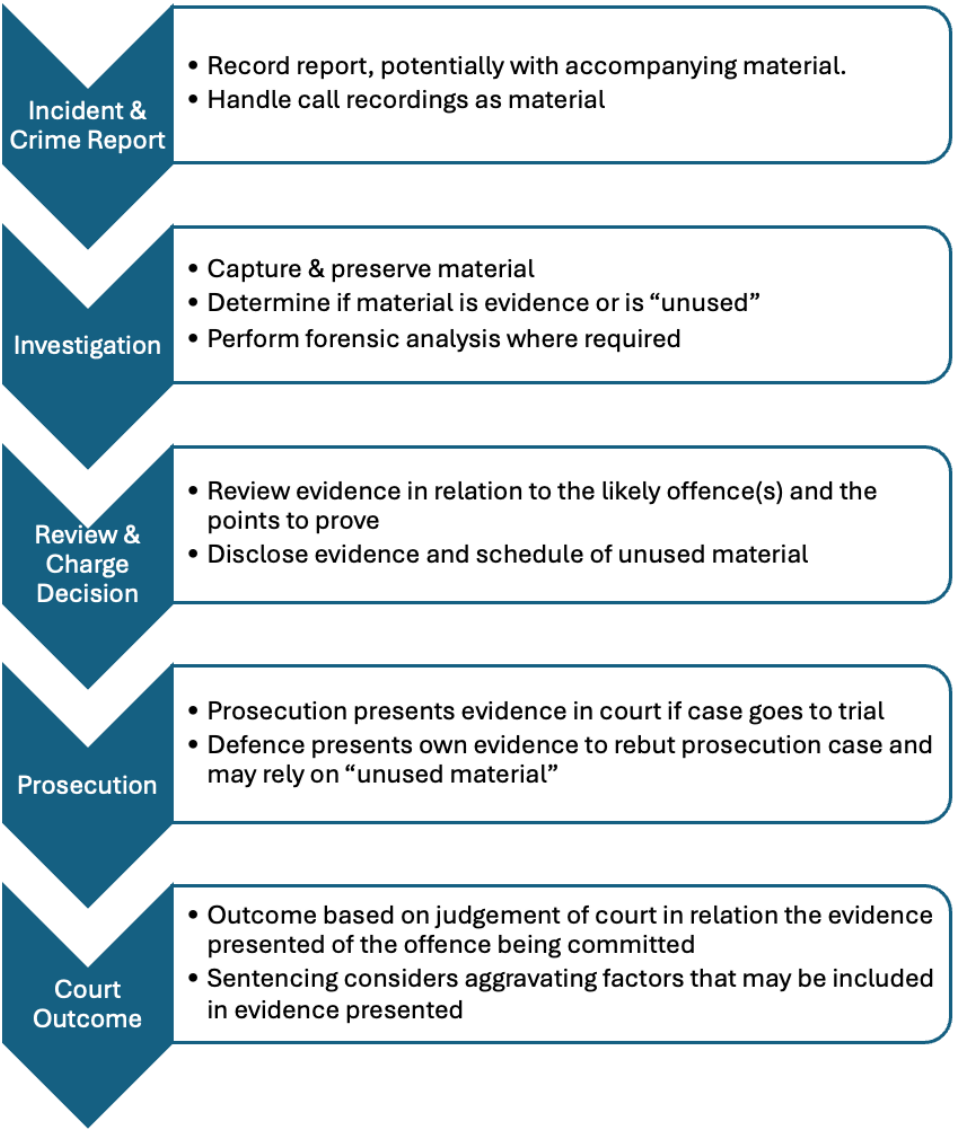



A thread running through is the concept of **relevance**. Evidence needs to be relevant to the offence and the points to prove and unused material needs to be determined to be relevant or not, largely in order to limit the scope of potential material. The relevancy test is therefore one of the key considerations. Similarly in the context of safety systems, evidence needs to be relevant to the system context and the safety case.

The concept of "Unused Material" in the AGO guidance, applies where there may be material that can potentially be relevant but potentially (subject to scrutiny) support the defence and/or undermines the prosecution case. Equally in safety cases, there may be a counter argument where existence of such material could be pertinent.

## A Lifecycle Perspective

Across the lifecycle of investigation and prosecution, evidence is handled and shared with the relevant participants. This is illustrated below in a much-simplified way.



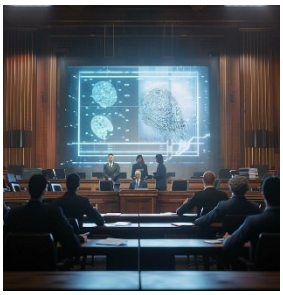
Handling material securely and providing transparency across the parties is a fundamental principle. Continuity and provenance need to be proven and documented as much as the case during the proceedings.

Accidental or even malicious tampering has been known to happen and current practices do the utmost to avoid these aspects from occurring. Information Confidentiality, Integrity and Availability (the security principles – CIA) are foundational principles.

**"Continuity and provenance need to be proven and documented as much as the case during the proceedings"**

Again the rules and regulations are embodied in the previously mentioned legislation (PACE, CPIA) and guidelines (AGO Disclosure) with relevant training especially for those roles with statutory responsibilities.

## Forensic Capabilities



International standards are particularly relevant in forensic analysis, with ISO 17025 regarded as the benchmark for accreditation of forensic laboratories and associated capabilities. Forensics covers a range of scientifically established techniques such as fingerprint analysis, DNA profiling and ballistics.

Locard's Exchange Principle holds that the perpetrator of a crime will bring something to the crime scene and will leave with something from it ("Every contact leaves a trace"). In our criminal damage example, we might therefore expect to have fingerprint and footprint analysis undertaken by an accredited laboratory and matched against relevant national, or potentially, international databases. Or else, against a print taken from the suspect once identified. Over the past few years, increasing use is made of mobile phone cell site analysis, images and video. These are time consuming and resource intensive, so typically the well-established techniques are used first.

Each of these techniques is typically developed in research environments and accuracy established using sophisticated analytical methods to build confidence and/or respond to challenges.

## Impact of Novel Technologies

Keeping up with the bad guys is a continual process of response and counter-response. The nature of evidence and the techniques used across the lifecycle need to keep pace. Some of the forensic technologies are mentioned earlier and there are many developments in the area of digital forensics. The increase in data and need for data quality management in policing was the topic of a paper at the SSS'18 Symposium [3]. The key points of that paper related to the increasing importance of data quality as the police and partner agencies move to more digital ways of working. Neglecting this aspect can have major consequences for people who might be wrongly pursued or prosecuted.



As the amount and complexity of material acquired increases, the resource implications rise. Hence the use of Artificial Intelligence & Machine Learning (AI/ML) is finding its way into material handling and analysis. Even so, the basics of establishing continuity and applying CIA apply.

Use of AI in fingerprint analysis is already with us, to assist with detecting ridge details and identify minutiae, which form the basis for matching. Policing is proceeding with caution in order to build confidence and establish trustworthiness.

Concerns about facial recognition prompted an in-depth analysis of operational technologies. In the report "Facial Recognition Technology In Law Enforcement Equitability Study" [4] the National Physical Laboratory (NPL) team developed a study methodology and conducted a controlled data capture, followed by statistical analysis to establish accuracy in real-world situations. This was also based on existing standards for biometric testing and reporting (ISO 19795 parts 1 & 2). The outcome of this work provides guidance for operational users in terms of setting thresholds that are needed in the trade-off between sensitivity & selectivity.




The theme of trustworthiness is a thread through numerous areas of analysis particularly as we all contend with increasing complexity of technologies, interconnected systems and the sheer volume of data that is part of modern life.

**"Trustworthiness is a product of the people, systems and processes within organisations…"**

Trustworthiness has been promoted by the UK Statistics Authority as part of their Code of Practice [5]. In the context of national statistics the other principles of Quality and Value apply. Trustworthiness in this area is defined as "… a product of the people, systems and processes within organisations that enable and support the production of statistics and data." Adopting and adapting this for data-centric technologies such as AI is already gaining traction. Standards and Codes of Practice for AI are introducing "Trustworthiness".

As a principle for evidence management Trustworthiness could be useful in both the safety and policing communities. Sitting alongside other key considerations such as relevance and quality. Potential adaptation of the underlying practices set out in [5] would be a useful activity. For example, can the practices written for management of statistics be easily translated into management of evidence, are any redundant and are any additional practices needed?

## Alignment of Evidence Management across Domains

Having set out in broad terms how material and evidence operates across policing we can turn our attention to where it is useful to align with safety engineering use of evidence. In safety, we look at evidence to support the development of a safety case and employ evidence to underpin the safety argument. If it all goes wrong, evidence is gathered during an investigation into an incident or accident.

That's a different starting point from evidence in policing, but broadly similar principles apply. Furthermore in those unfortunate circumstances where an incident might be dealt with through the criminal procedures, the evidential standards need to be fully aligned. Very briefly here are a few considerations to indicate the areas worth exploring further.

**Parallels:**

a) In policing the key is gathering evidence to support the "point to prove" – just like we gather evidence to support a safety argument or the "claim"
b) There is legislation, standards and guidance governing processes
c) If processes are not followed then the prosecution can be challenged and might fail. Similarly, flawed evidence in a safety case may result in failure to be certified or ultimately manifest as a higher risk of accident
d) There are process lifecycles where evidence can be pertinent and collected at each stage and each has an important part to play in the overall case. Accurate record keeping, provenance and demonstrating correct handling is important.

**Differences:**

a) Novel evidential techniques are being developed all the time but this is perhaps less so in the safety domain that relies on tried and tested techniques. This is illustrated in the challenges of assuring AI – we don't have the techniques to do that reliably in safety at the moment. So we have novelty in systems but not novelty in techniques.
b) Laboratory or testing facilities, at least in the UK, typically have quality management systems and meet the essential requirements around accreditation of capabilities and facilities plus equipment calibration standards apply across domains. In policing however, the specialist support and scrutiny of capabilities has been applied since 2007 through the Forensic Science Regulator with statutory powers provided through the Forensic Science Regulator Act 2021.
c) Safety doesn't usually make much of evidence not selected as relevant although we are starting to see more dialectic arguments where the safety case deliberately seeks counterevidence for the claim. There is less formality about whether evidence is relevant (perhaps equivalent to a disclosure test) and there is scope for confirmation bias if data is ignored that doesn't fit the argument.
d) Maybe the most significant differences are in the people aspects. In UK policing there are statutory responsibilities and defined roles such as investigators and disclosure officers, holding the authorities to account in a very testable manner. Whilst in safety engineering, there is a less prescribed set of roles and responsibilities with some aspects more open to interpretation.
e) Prosecutions need to be understandable to the many (ultimately a jury of lay-people).

In safety, the safety case at times is understandable only by the few and requires specialist knowledge. Although, if an incident with safety implications results in a public enquiry or other proceedings, the understanding of evidential material will equally apply.

**Potential improvements:**

Some ideas based on practices within policing for potential improvements within the safety community are as follows:

- There may be opportunities to better define a taxonomy in safety engineering adopting whatever makes sense to do so from policing or indeed other domains. Establishing and maintaining a taxonomy of evidence and related matters can help ensure a common understanding, language and interpretation
- Consider something akin to a disclosure test and relevancy that determines if information can be safely set aside or explicitly regarded as evidence but also have a means of assessing if previously discarded evidence becomes relevant in the future.
- Ultimately it may be useful to adopt the mindset of an investigator or prosecutor. Considering if there was an accident and a resulting prosecution, how would the evidence/safety case stack up to form a defence? Are there special things we should be doing to help support such an eventuality? Perhaps even playing again to the ideas of storytelling and having a narrative in safety cases. Should safety cases be positioned so they are understandable to lay people like a jury?

Practitioners in both domains are encouraged to take time to understand the similarities and differences and hopefully this article will prompt some useful further discussion of mutual benefit.

## References


[1] Attorney General's Office "Attorney General's Guidelines on Disclosure " Gov.UK, 29 February 2024

[2] Various legislation for illustration: CDA 1971, PACE 1984, CPIA, FSR Act https://www.legislation.gov.uk/

[3] Data in Police & Criminal Justice Systems: How data errors could lead to harm to innocent citizens, Bill Blackburn, Paul Hampton, Mike Parsons SSS'18

[4] Facial Recognition Technology in Law Enforcement Equitability Study Final Report, Dr Tony Mansfield, NPL Data Science Department, March 2023 https://science.police.uk/site/assets/files/3396/frt-equitability-study_mar2023.pdf

[5] Code of Practice for Statistics Ensuring official statistics serve the public, UK Statistics Authority Office for Statistics Regulation Edition 2.1 Code revised 05 May 2022



Image Attribution:
leading image: AI generated by MidJourney
fingerprints: AI generated by MidJourney
Facial recognition: ID 220039699 © Andrey Popov | Dreamstime.com



**Bill Blackburn, Process Renewal Group Limited and SCSC Member**

Bill has a background in mechanical engineering and has worked across aerospace, manufacturing, defence and other high integrity sectors for over 40 years. He is presently engaged in advisory roles with clients in UK policing and capability development in Europe. He has extensive experience in business and information architecture, with a focus on process and data management.

# Is there something hidden in the code?



**Gavin Wilsher describes an example of where undesirable code appears to have been deliberately inserted into a train control system rendering it inoperative after maintenance at a 3rd-party site. He explores whether the safety community should do more to ensure these types of issue do not find their way into safety-related / critical software whether deliberate, inadvertent or through the increasing use of Artificial Intelligence.**

This article was prompted by a well-publicised incident that occurred on a particular train type in Europe in 2023 [1]. Readers of this newsletter may already be familiar with the story. There are certain similarities between this story and the well-publicised issue involving vehicles with diesel engines having hidden code to ensure that emissions tests could be passed.

It was reported on the RailTech website in January 2024 that the train manufacturer was taking legal action against the maintenance company [2] although it is not clear what progress has been made. All information in this article directly related to the incident is in the public domain.

The intention here is not to analyse that particular incident but rather to look under the surface and wonder if there is a more endemic issue in the railway sector (and potentially other safety-critical sectors) that should prompt more safety interest. Hopefully, this will serve to get a conversation going (even if it is simply to tell me that I might be worrying about nothing).

This article presents a summary of the incident I refer to above, but is only informed by it and not truly about it. It is really about things that can be hidden in "the code" and whether we should be concerned. However, I think it is important to understand what appears to have happened, how it was investigated and subsequently resolved before concluding whether there is a wider problem.

Nobody can be sure of the motivations for the train manufacturer to do what they appear to have done, although some have speculated that this was done as a direct result of them not winning the maintenance contract in a competitive tendering situation. Looking at the issue superficially, it is easy to reach this conclusion, but the fact is that nobody outside the company can be sure. It is certainly not the purpose of the article to wonder why they did it.

The following few paragraphs summarise what was reported to have happened.

## The Issue




An Impuls 45WE train, manufactured by Polish company Newag, that was undergoing mandatory, scheduled maintenance by a recognised maintenance organisation unexpectedly failed to run once the maintenance had been completed. The train appeared to be fine, it just failed to run.

The Lower Silesian Railway in Poland runs eleven of these types of trains and, according to the maintenance schedule, another one was due to be sent for maintenance while the first – instead of returning to passenger service – was still sitting in an apparently failed state in the workshop.

The second train then underwent the same maintenance as the first with the same resulting outcome – it failed to run following completion of the maintenance. The maintenance team continued to try to get the first train running, without success. It is reported that the manufacturer was approached for assistance but they declined to help. If this is indeed the case, it raises questions about the management of contracts, which could have been written such that there were obligations placed on the manufacturer to support maintenance or other technical issues.

This scenario continued to play out on further trains, including one at another maintenance depot belonging to the maintenance company.

The maintenance organisation was conducting the maintenance in accordance with the manufacturer's instructions, provided in a 20,000 page manual. At the completion of the maintenance it was expected that the train would run and be returned to passenger service. However, the inverters failed to supply voltage to the motors and no one had any idea why this was happening. Maintenance technicians conducted deep investigations but failed to find the cause.

The problem came to the attention of the press because timetables were affected and replacement trains were needed in order to maintain some level of service.

The maintenance company was incurring substantial penalties by this time and was anxious to get to the bottom of the problem.

At some point they reached out to a group of hackers to see whether they might be able to help get into the train's computers to understand what was happening.

Many months of analysis and reverse engineering ensued and they eventually uncovered some rather interesting conditions written into the software code of the trains. What they discovered was the GPS co-ordinates of maintenance depots hidden in the code. Furthermore, there was a condition written into the code that prevented the train from running if it spent more than 10 days at one of these locations. They also found the GPS co-ordinates of a depot belonging to the train manufacturer, but this had a different set of conditions associated with it.

The whole thing was presented at the Chaos Communication Congress in December 2023. A video of the presentation is available in [3]. It is well worth watching.

## So what?

There is no suggestion in any of the commentary about this issue that there were any safety implications arising from it. The immediate impacts were potentially to the maintenance organisation's reputation and their bottom line. It is also the case that passengers seem to have been affected by the timetable changes and overcrowding on the smaller, replacement trains that needed to be deployed. Having said that, overcrowding can have a minor safety impact.

**"The real concern is that manufacturers can hide such information within code that is presumably subject to internal code reviews, without challenge"**

However, the real concern is that manufacturers can hide such information within code that is presumably subject to internal code reviews, without challenge. The hidden information was not highlighted in the extensive maintenance manual, despite it being directly related to maintenance, so the documentation could be seen to be incomplete i.e. why wasn't there a warning in the maintenance manual that after 10 days in the depot the train may fail to run? The information related to the depot belonging to the train manufacturer was presumably in the code so that the conditions could be tested. If that is the case then it can be assumed that if the software was subjected to internal Verification and Validation (V&V), that it passed without anyone requesting the removal of what appears to be information that is unrelated to correct functioning of the software. Indeed, effort may have been expended testing that the conditions associated with this additional information actually functioned as intended. That is probably less concerning because at least it would indicate some rigour was applied to making sure this apparently unnecessary information performed as intended.

In the train scenario described above there is no direct relationship with safety. However, what it illustrates is just how easily seemingly unnecessary information can be hidden away in the code. What if something like this was added to the Safety Integrity Level (SIL) 4 <insert any high integrity safety-critical software from any sector here> code? Can we assume that, provided it passes internal V&V then the software will be deployed onto the market?

Under specific conditions the hidden information may lead to an error condition that causes a safety hazard. Failure to find these conditions during code analysis and test does not prove they don't exist, all it does is prove that the testing failed to find them.

Questions have been raised on this matter on various online forums along the following lines:

- How was the train manufacturer able to do this?
- Why isn't there an independent overseer auditing and certifying that what the computer code does is exactly as documented?
- Why was the train manufacturer not obliged to send source code to the railway operator?

These sort of questions do seem to point to the standards and regulations and what they mandate (or do not mandate) in terms of independent review.

Taking this issue at face value, I have become convinced of two things:

- We should be concerned that this sort of problem may end up crossing over into high integrity software intended for safety applications at some point
- This is probably not the first time something of this nature has happened, it is just that this time, it has been made very public

Set aside for a moment that the issue that prompted me to write this article was created by humans acting in a deliberate manner; there is currently growing emphasis on using so called Artificial Intelligence (AI) in software engineering as a 'co-pilot' and in writing test cases. This drives me to wonder whether this might start to become much more of a problem, especially given the current propensity of AI to 'hallucinate' i.e. make things up. The AI may insert random artefacts into the code that could, under certain conditions, cause a safety hazard but not be detected during code analysis and test, particularly if the same AI engine is used to manage the testing or analyse the code.

I am convinced that this is something we need to start to think much more carefully about. There has already been significant thought about this by the SCSC's Safe Autonomous Systems Working Group.

**"following standards like IEC61508 and applying the techniques and measures defined therein should be sufficient. But is it?"**

## What could we do about it?

So, do we need a more rigorous and complete approach to independent (third-party) code review? We could argue that following standards like IEC61508 [4] and applying the techniques and measures defined therein should be sufficient. But is it? Software safety standards often mention redundant code but perhaps they also need to consider the assurance of the absence of non-required or even adversarial code.

I have experienced a situation where SIL4 software used in a train control system was developed correctly in accordance with EN50128 [5], did exactly what it was supposed to do, passed all the module and system level tests but still produced an unsafe failure condition when deployed onto the railway. Fortunately, this was detected by a vigilant signaller before any harm was done.

The particular condition in the software that caused this issue could very easily have been the result of the addition of geolocation data such as that described in the Impuls 45WE scenario. With the benefit of hindsight, this now adds to my belief that we need to do more than we currently are doing to protect against this.

I would like to see an independent, international working group set up, comprising representatives from commercial organisations and the regulators, to explore practical ways to deal with potentially unnecessary information being hidden in the code. Although all of the suggestions I make in this article involve different approaches to independent code review, the working group should also consider whether there are other approaches that could be used i.e. activities other than code review.

The working group could then publish informal guidance in the form of a community standard and work towards getting the guidance introduced formally into international standards to make it mandatory.

Some thoughts about the type of questions the working group could address are presented in the following paragraphs.

Should the independent assessor be responsible for a full, independent code review? That sounds expensive and would depend upon the assessor having a comprehensive understanding of not only software engineering but the multiplicity of programming languages that may be used in safety-critical applications.

Perhaps the relevant regulatory body should be responsible for conducting sample based code reviews e.g. select software modules and perform a full review of the code? These samples could be taken using a risk-based approach to ensure that targeted effort is applied where the consequences of apparently unnecessary information being hidden in the code are the greatest. The temptation might be to use the highest SIL to select the samples and this sort of approach would be most useful where there are multiple SIL 4 software modules to select samples from.




Perhaps the end user (train operator, airline, power station operator etc.) should be responsible for undertaking code reviews? But surely they depend upon the independent assessor, the regulator and even the manufacturer themselves to assure them that the system is safe enough to put into use?

It would be easy to jump to the conclusion that the answer to whether more third-party reviews should happen is 'yes' but that may be because, in the world of safety-critical systems, we tend to be risk averse and more checking tends to make us feel more comfortable.

So, what about the standards and regulations? Perhaps they should be updated to mandate fully independent review of such code if the code can be accessed by any safety-related applications (either intentionally or otherwise).

Thinking about my concerns regarding the use of AI in software engineering, the company I work for has very strict guidelines about how, when and to what extent AI can be deployed. I have no idea whether we are unique or whether that is a standard approach. However, I do think that a cross-industry standard should be mandated that sets out certain rules regarding the application of AI to software engineering in safety-critical fields. I do not believe that the International Standard on AI, ISO/IEC 22989 [6] does this for us. Perhaps an update to IEC61508 is required. Currently it states (in IEC61508-3) that AI is not recommended at SIL 2 and above. AI has moved on a lot since the current version was published in 2010.

## Concluding Remarks

This article has sought to question whether, in light of the train maintenance issue described above, the safety community (us) need to be considering more rigorous third-party reviews of software code. I have also used it to raise my concerns about the growing use of AI in software engineering and whether there needs to be a standard that covers its application in safety-related / critical software.

In writing this article, I did not set out to answer these questions (it would be arrogant of me to suggest that I have all the answers) but to raise some awareness and pose some questions that might start a conversation that does indeed lead to answers.

I often provide independent safety assessment services as part of my current role and this issue has certainly impacted the way I think about assessing the software and has prompted me to ask more probing questions about code reviews and what they have considered.

## References


[1] https://badcyber.com/dieselgate-but-for-trains-some-heavyweight-hardware-hacking, accessed April 2024

[2] https://www.railtech.com/policy/2024/01/03/poland-hackers-used-to-unblock-trains, accessed April 2024

[3] https://media.ccc.de/v/37c3-12142-breaking_drm_in_polish_trains, accessed April 2024

[4] International Electrotechnical Commission (IEC), "IEC61508:2010, Functional safety of electrical/electronic/programmable electronic safety-related systems"

[5] CENELEC "EN50128 Railway applications - Communication, signalling and processing systems - Software for railway control and protection systems" 2011+A2:2020 (superseded by EN50716:2023)

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



Image attribution:
top image: AI generated: MidJourney
Newag Impuls train at station in Wegliniec: ID 171096643 © Oxygen64 | Dreamstime.com
Code review: ID 49028202 © Rasmus Ursem | Dreamstime.com



**Gavin Wilsher**

Gavin Wilsher is a systems and safety engineer with over 25 years' experience with safety-critical systems across many regulated industry sectors. This experience includes safety engineering and independent safety assessment in the civil nuclear, military aerospace and railway signalling sectors.



The author retains copyright of this work.

# AI Safety: Navigating the Expanding Landscape of Potential Harms



**Ibrahim Habli and John McDermid look at how the emergence of Artificial Intelligence (AI) is adding pressure on the safety community to rethink the boundaries of safety engineering. The rise of AI is exposing a wider spectrum of harms, with significant moral, social and economic impacts. Can the safety-critical community realistically expand the scope of safety to address these intangible yet impactful AI-induced harms?**

Traditional safety management focuses on preventing accidental harm. But given that absolute prevention is rarely possible, the focus of safety activities tends to shift to risk, i.e. the likelihood and severity of harm. The kinds of harm of particular concern are typically physical in nature, directed to humans, as well as to property and the environment. The increasing use of Artificial Intelligence (AI), whether in safety-critical applications or not, extends the scope further. It raises concerns about the potential of the technology to cause other significant harms [1]: moral, social, political, and economic [2]. This has led to growing pressure and expectations for "safety" measures to address these wider categories of harm [3], and the term "AI Safety" is often used to encompass managing the risks from AI for these wider categories of harm. Whilst such concerns might seem to fall naturally within the remit of safety engineering, can the safety-critical community realistically expand the scope of safety to address these intangible yet impactful AI-induced harms? Is such an expansive approach feasible, and if so, what strategies will be necessary to navigate this vast and uncharted territory?

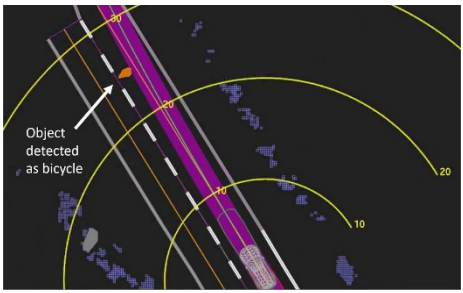
## From Physical to Psychological and Environmental Harm

Safety-critical industries such as transport, defence and manufacturing have traditionally placed unique emphasis on eliminating or reducing the risk of physical harm caused by sociotechnical systems on humans and on property. Death, injury, and damage to physical artefacts have been the core categories of concern, supported by a rigorous ecosystem of standards, guidelines, and best practices, which have gradually been evolved by the relevant subject-matter and professional communities.

Further, in the last two decades, we have witnessed a critical and growing recognition of the psychological dimensions of harm. In particular, the development of highly interactive digital technologies and software-intensive automation systems, while enhancing performance and physical safety, has introduced or shed new lights on psychological risk factors. The potential for anxiety, stress and trauma, particularly in high-pressure environments and complex human-machine interactions, is now better understood and proactively analysed [4]. Examples include air traffic controllers or clinicians experiencing fatigue and burnout due to complex workflows or understaffed settings.

This historically narrow scoping of harm has undergone, albeit belatedly, an expansion in recent years, leading to the inclusion of environmental impacts within the scope of safety management [5]. Driven by the climate crisis, issues such as ecological damage and pollution are now better recognised as potential consequences of activity in safety-critical industries, particularly in the transport and energy sectors.

AI systems, mostly operating in 'testing' phases or 'advisory' roles, have been linked to actual physical, psychological and environmental harms. For example, in 2018, an Uber test vehicle, operating in a self-driving mode, hit and killed Elaine Herzberg in Tempe, Arizona [6]. This fatal accident illustrates how AI-enabled features could contribute, amongst other key factors, to human harm. The Automated Driving System (ADS), which relied on machine learning, inaccurately classified Herzberg and her trajectory.




Automated crash avoidance, a function of the ADS, was not possible by the time the system predicted that a crash was imminent. While the deficiencies in the design and risk assessment of ADS were highlighted as a contributory factor, the National Transportation Safety Board (NTSB) investigation identified the *'failure of the vehicle operator to monitor the driving environment and the operation of the automated driving system'* as the '*probable cause of the crash'.* The complexity of the human-machine interaction and Uber's safety culture, increased the risk of 'automation complacency', whereby the vehicle operator over-relied on the capability of ADS.

## Intent Matters

In safety management, the focus is primarily on *accidental* harm, arising from errors or mistakes. Any harmful consequences are unintended, in contrast to *deliberate* harm caused by malicious actors. This has formed a basis for the conceptual difference between safety (freedom from accidental harm) and security (protection from malicious harm). It has led to

different specialised methods such as hazard analysis in safety engineering and attack trees in security engineering. However, recognising this difference alone is counterproductive [7].

In increasingly interconnected systems and digitised services, safety hazards can emerge from security vulnerabilities. An unsecured software-based system, such as a Supervisory Control and Data Acquisition (SCADA) system in a power plant, might expose operational networks to external malicious attacks, potentially triggering safety-critical industrial accidents or incidents [8]. This reinforces the need for an integrated approach that transcends the accidental and intentional gap, perhaps with 'risk' as the integrating concept. We now accept that safety hazard analysis must proactively consider security threats as potential sources of harm, and cybersecurity methods must analyse the safety implications of any additional security controls.

The boundaries between safety and security are even more blurred for AI systems [9]. With their reliance on large and often interconnected data sources, deep neural networks for example are particularly prone to adversarial attacks. In particular, perception and vision functions, implemented by these models, are vulnerable to manipulation by malicious actors, e.g. fooling the networks by placing small stickers on traffic signs [10], leading autonomous cars to misclassify stop signs and increasing the risk of collision.

## Extending AI Safety to Societal Harms

Policy and regulatory documents already emphasise the potential benefits and dangers of the use of AI. This is evident in, for example, the UK's pro-innovation regulatory approach to AI [11]. As outlined in the Policy Paper published in March 2023, this approach recognises the wide range of AI-induced harms, from the potential of the technology to '*damage our physical and mental health*' to its capacity to '*infringe on the privacy of individuals and undermine human rights*'.

Increasingly, the latter two kinds of societal harms, as well as other broader kinds of ethical and political concerns, feature under the umbrella of AI safety [12, 13]. These include algorithmic discrimination in recruitment and in the justice system, racial bias in accessing health and social care services, targeted generation of misinformation by malicious actors, sudden shifts in the job market and overdependence on a handful of powerful AI companies. The concerns go beyond the individual instances and extend to troubling society-wide patterns that could weaken long-established democratic, social and regulatory institutions and norms. These sources of harm are not new. However, the scale and complexity of AI systems exacerbate and entrench these systemic issues. Further, the opaque nature of AI systems, specifically those incorporating deep neural networks, which may be trained and tested using imbalanced and biassed data, are inaccessible and inscrutable by those directly affected by their deployment (e.g. patients or pedestrians) or their representatives (e.g. regulators or advocacy groups). This leads to significant legal, social and ethical responsibility gaps [14].

**"The opaque nature of AI systems, specifically those incorporating deep neural networks, which may be trained and tested using imbalanced and biassed data, are inaccessible and inscrutable by those directly affected by their deployment"**

**"An AI model might be deemed 'safe enough' to use, but its development might raise significant ethical and even legal concerns".**

It is also important to consider the concerns in the AI development process itself, i.e. 'upstream harms' [15]. These include unfair working conditions for those collecting and labelling machine learning datasets, the environmental footprint of training and running foundational AI models, and privacy violation around the use of publicly available data. That is, an AI model might be deemed 'safe enough' to use, but its development might raise significant ethical and even legal concerns.

The rapidly growing landscape of AI harms raises a critical question: what is the appropriate scope for AI safety? To illustrate the potential for harm across various contexts, let's consider these realistic scenarios:

A. A generative AI system that spreads fake news, eroding confidence in independent regulators
B. An AI-based sentencing system that produces racially-biassed recommendations
C. An autonomous car that misclassifies stop signs and leads to collision
D. An AI-based clinical-decision support system that weakens trust in human clinical judgement
E. An AI-based health screening system that is at least as performant as current non-AI services but underperforms for patients with dark skin
F. Generative AI content promoted on social media that undermines confidence in election results

These scenarios highlight how AI can cause harm beyond the traditional focus on physical or psychological impacts. There are several approaches to framing the scope of AI safety harms:

1. **Broad Scope**: This approach addresses ethical, legal, social, political, and economic harms (covers all scenarios)
2. **Traditional Focus**: This approach prioritises physical and psychological harm to humans, along with physical damage to property and the environment (only scenario (C) applies)
3. **Refined Focus**: This approach expands option 2 by including additional harms from option 1 that refine existing categories in as far as the nature of the final harmful outcome is physical, psychological or environmental in nature (scenario (E) joins (C))
4. **Direct Causation**: This approach expands on option 3 by including additional harms from option 1 as direct causal factors (scenario (D) joins (C) and (E))
5. **Comprehensive Approach:** This approach builds on option 4 by considering additional harms from option 1 as contextual or indirect causal factors (scenario (A) joins (C), (D), and (E). Scenarios (B) and (F) might also be relevant, although their direct impact is harder to establish)

Option 1, while offering a holistic view, might overwhelm current safety methodologies. Option 2 is too narrow, neglecting crucial aspects. Options 3-5 offer a spectrum of possibilities, with Option 5 seeming the most promising. It maintains a focus on core harms (physical, psychological, environmental) while explicitly addressing equity in safety. Additionally, it considers other potential harms (such as misinformation) as contributing factors, striking a balance between scope and feasibility.

## Where do we go from here?

The answer does not lie in artificial boundaries or fragmented efforts. Fundamental safety norms endure: starting early, adopting a whole-system mindset, and maintaining safety through-life remain unchallenged by AI. AI or not, physical and psychological harm to humans and physical damage to property and the environment will rightly remain the focus of safety activities. However, an open and inclusive debate is unfolding on how to integrate other AI-induced harms, especially ethical and legal concerns about bias and discrimination. This should involve methods for assessing algorithmic fairness and meaningfully incorporating diverse stakeholders throughout the AI development process. Whether this is reflected in an expanded scope for safety engineering, or in a new all-encompassing "AI Safety" interdisciplinary profession remains to be seen. Whichever emerges, it is clear that the next generation of professionals will need broader, multi- disciplinary knowledge – something we are seeking to foster in our UKRI Centre for Doctoral Training in AI Safety (SAINTS) [16].

The aim of SAINTS is to train a cohort of 60 PhD students with the research expertise and skills necessary to ensure that the benefits of AI-enabled systems are realised without introducing harm as the systems and their environments evolve. Research will be focused on the lifelong safety assurance of increasingly autonomous AI systems in dynamic and uncertain contexts, building on methodologies and concepts in disciplines spanning AI, safety science, philosophy, law and the social sciences. SAINTS will bring together students and partners from a diversity of backgrounds and sectors to deliver a new generation of experts who make leading contributions to AI safety.

## References


[1] Zoe Porter, Ibrahim Habli, John McDermid, and Marten Kaas. "A principles-based ethics assurance argument pattern for AI and autonomous systems." AI and Ethics (2023): 1-24.

[2] Jess Whittlestone, Rune Nyrup, Anna Alexandrova, Kanta Dihal, and Stephen Cave. "Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research." London: Nuffield Foundation (2019).

[3] Department for Science, Innovation and Technology. "Frontier AI: capabilities and risks – discussion paper." https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper (October 2023).

[4] Sidney Dekker. Second victim: error, guilt, trauma, and resilience. CRC press, 2013.

[5] John Alexander McDermid, Simon Burton and Zoe Porter. "Safe, ethical and sustainable: framing the argument." In The Future of Safe Systems: Proceedings of the 31st Safety-Critical Systems Symposium, 2023, pp. 297-316. Safety Critical Systems Club, 2023.

[6] National Transportation Safety Board. "Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe, Arizona March 18, 2018", https://www.ntsb.gov/investigations/accidentreports/reports/har1903.pdf. (2019)

[7] Robin Bloomfield, Kateryna Netkachova and Robert Stroud. "Security-informed safety: if it's not

secure, it's not safe." In Software Engineering for Resilient Systems: 5th International Workshop, SERENE 2013, Kiev, Ukraine, October 3-4, 2013. Proceedings 5, pp. 17-32. Springer Berlin Heidelberg, 2013.

[8] David Kushner. "The real story of stuxnet." ieee Spectrum 50, no. 3 (2013): 48-53.

[9] Department for Science, Innovation and Technology. "The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023". https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

[10] Nir Morgulis, Alexander Kreines, Shachar Mendelowitz and Yuval Weisglass. "Fooling a real car with adversarial traffic signs." arXiv preprint arXiv:1907.00374 (2019).

[11] Department for Science, Innovation and Technology. "Policy paper A pro-innovation approach to AI regulation" https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper (3 August 2023)

[12] Emily M Bender, Timnit Gebru, Angelina McMillan-Major and Shmargaret Shmitchell. "On the dangers of stochastic parrots: Can language models be too big? 🦜." In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, pp. 610-623. 2021.

[13] Ada Lovelace Institute. "Regulating AI in the UK", https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk (July 2023)

[14] Simon Burton, Ibrahim Habli, Tom Lawton, John McDermid, Phillip Morgan and Zoe Porter. "Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective." Artificial Intelligence 279 (2020): 103201.

[15] Ada Lovelace Institute. "Regulating AI in the UK", https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk (July 2023)

[16] UKRI AI Centre for Doctoral Training in Lifelong Safety Assurance of AI-enabled Autonomous Systems (SAINTS): https://www.york.ac.uk/study/postgraduate-research/centres-doctoral-training/safe-ai-training

Imagine Attribution:
leading image: AI generate by Midjourney
Tempe accident: Public Domain: https://commons.wikimedia.org/wiki/File:Tempe_prelim_figure_2_(41583363594).png

**Professor Ibrahim Habli, Research Director, Centre for Assuring Autonomy, University of York**

A leading expert in design and assurance of safety-critical systems, Ibrahim works closely with industry and regulators and has made significant contributions in dynamic safety cases and safe and ethical use of AI. Driven by a passion for interdisciplinary safety research and team science, he has led major, complex research activities, with responsibility for collaboration with organisations like the NHS and Jaguar Land Rover. He also directs the UKRI Centre for Doctoral Training in AI Safety (SAINTS).

**Professor John McDermid, Director, Centre for Assuring Autonomy, University of York**

An internationally recognised speaker and expert in system, safety and software engineering, his role at the centre focuses on safety of robotics and autonomous systems. He has acted as an advisor to government and industry for several decades, working in several organisations as a Non-Executive Director, including the Health and Safety Executive (HSE), and as an Advisory Board Member of organisations such as the Ministry of Defence, Rolls-Royce and the Fraunhofer Institute in Kaiserslautern. In 2002 he became a Fellow of the Royal Academy of Engineering and was awarded an OBE in 2010.

The authors retain copyright of this article.

# Letters to the Editor



Dear Paul,

I am writing to you in response to Gavin McCall's presentation at SSS'24 [1]. I was disappointed that someone developing software for a safety application has so little confidence in his own team's ability to develop high integrity software and so much confidence that strangers could do better.

Gavin's presentation started with an example where a developer was faced with the choice between developing 500,000 Source Lines Of Code (SLOC) of bespoke software, and using open-source software instead. It is very likely that the open-source software will contain many features that are not needed in this particular application. Supposing the open-source software contains 2 million SLOC, it follows that the open-source software would need to exhibit a defect rate (expressed as defects per thousand lines of code) four times better than the bespoke software to result in the same number of defects overall.

In my experience, most safety applications contain far less than 500,000 SLOC. It is true that there has been an exponential increase in the size of many software applications. Software productivity hasn't improved exponentially, neither have team sizes. Most of the increase in code size has come from reusing third party software, including Commercial Off The Shelf (COTS) and open-source software. While this can be a means of delivering software quickly when the software only needs to be "good enough", it is not an effective means of developing high integrity software. I am reminded of Blaise Pascal, who wrote "I have made this [letter] longer than usual because I have not had time to make it shorter".

Gavin's presentation used the Linux kernel as the only example of open-source software. I agree that the Linux kernel appears to be very reliable. I don't know how it compares with operating systems targeted specifically at safety applications, such as BlackBerry QNX, Green Hills Software INTEGRITY or DDC-I Deos. However, Linux is much more than just the kernel.

It doesn't follow that all Linux distributions are suitable for supporting safety applications. The SIL2LinuxMP project tried for three years to qualify Linux as a pre-existing software element following route 3S in IEC 61508-3, with the goal of providing a minimal run-time environment for safety-critical applications up to SIL 2 (See "Certifying Linux: Lessons Learned in Three Years of SIL2LinuxMP" [2]). The Linux Foundation's Enabling Linux in Safety Applications (ELISA) project is currently looking at how to make it easier for companies to build and certify Linux-based safety-critical applications, though they emphasise they are not producing a safe Linux distribution (see "ELISA - Advancing Open Source Safety-Critical Systems – ELISA" [3]).

The fact that the Linux kernel is good does not mean that all open-source software is good. You cannot extrapolate from the specific to the general. Most open-source software is of much lower quality than the Linux kernel. I was asked to assess an open-source software library developed by a major corporation. The open-source software library contained 381,512 SLOC of software, but only 5,291 SLOC of tests. This is typical of most open-source software developments, where the number of developers greatly exceeds the number of testers.

I have been privileged to have worked with some world-class programmers, including but not limited to Dave Thomas [4], Peter Amey [5] and Rod Chapman [6]. The software development process that I helped to develop at Praxis, called Correctness-by-Construction, was recognised by the US Department of Homeland Security as one of only four software development processes worldwide that can consistently deliver a defect rate of less than 1 defect per thousand lines of code (see "Processes to Produce Secure Software" [7]).

Open-source software is a not a silver bullet, nor is it a substitute for sound software engineering practices. The processes and techniques required to deliver high integrity software have been well-known for decades and applied by responsible software developers. The state of the art continues to advance, as described by Rod Chapman in his excellent talk at HISC 2023: "Automated Reasoning in and about the Cloud" [8].

Yours,

Eur Ing Dewi Daniels CEng FBCS

## References


[1] RAFIA - Using open-source software in an Automotive safety application - a case study, Gavin McCall, https://scsc.uk/re1007.39:1

[2] https://www.researchgate.net/publication/323445105_Certifying_Linux_Lessons_Learned_in_Three_Years_of_SIL2LinuxMP

[3] ELISA - Advancing Open Source Safety-Critical Systems – ELISA, https://elisa.tech/

[4] Dave Thomas (programmer) – Wikipedia, https://en.wikipedia.org/wiki/Dave_Thomas_(programmer)

[5] Peter Amey, https://www.sigada.org/Peter-Amey.pdf

[6] Rod Chapman https://www.linkedin.com/in/rod-chapman-7b60266

[7] Processes to Produce Secure Software, https://apothecaryshed.files.wordpress.com/2018/11/secure_software_process1.pdf

[8] Automated Reasoning in and about the Cloud, High Integrity Software Conference, https://www.his-conference.co.uk/session/automated-reasoning-in-and-about-the-cloud

# SSS'24 Event Report



**The Safety-Critical Systems Club's annual Symposium returned in February 2024 for another fully in-person event hosted at the Bristol Royal Marriott. Paul Hampton, the SCSC Newsletter Editor, provides highlights of the three-day event.**

Mike Parsons introduced the symposium and noted that this year's event was very well attended with 176 delegates. Mike said one of the key themes of the symposium was managing the risks around Artificial Intelligence (AI), and so amongst other AI-related talks, a special panel session was going to be held on the second day allowing delegates to pose AI-related questions to the panel of experts.

New for this symposium was the SCSC timeline (see page 43), and other features introduced last year such as the crossword, roving video interviews and a light-hearted take on last year's "Room 101" where delegates could, this time, vote on the most compelling narrative from a panel providing stories of events that influenced or even change their system safety world.

## Byzantine Generals Attack an Airbus A320



In the first key-note speech of the Symposium, Dewi Daniels expanded on a talk he had previously given in 2022, which discussed an Airbus A320 incident involving the failure of all three primary flight control computers on touchdown.

Airbus claimed from in-service data that no similar fault had occurred in over 44.3 million flight hours, but Dewi said that another incident did occur relatively soon after his previous talk. Dewi said that both incidents suffered from the same problem – "The Byzantine Generals Problem", first described by Leslie Lamport in his relatively unknown 1978 paper: "The implementation of Reliable Distributed Multiprocess Systems". This was reimagined as "The Byzantine Generals Problem" in 1982, which presented the ideas more in a story-telling narrative, and this has now had over 10,000 citation showing the power of storytelling in communicating ideas and concepts.

Dewi reprised the problem Byzantine General Problem:

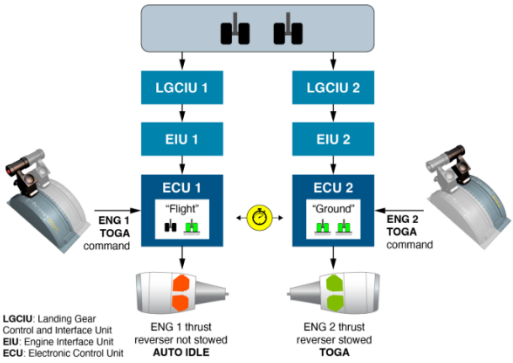
- A General in charge of several Lieutenants must get agreement on common actions amongst the lieutenants for the overall campaign to be successful
- However, one of the commanders is unreliable and so may send misleading messages to the other commanders to sabotage the campaign

Dewi said this is effectively a form of 2/3 voting system. Lamport pointed out that if input is faulty then it could confuse the voting systems. Faults may also arise if the processors read the input at different times.

Dewi showed how this mapped to the Airbus Flight control architecture where monitor system were employed that would switch off a command lane if errors were detected. In the original Airbus incident, cross winds and bouncing on landing caused the 'weight on wheels' sensor readings to fluctuate resulting in different values being read by the monitors. Dewi noted that there were no system faults here, it was simply that the readings were not in synchronisation. In less than a second, all three primary computers had shutdown and the pilots were only able to stop aircraft through manual braking.

Dewi then discussed the 2022 incident where an A320 was landing at Copenhagen during a cross wind and he shared the Airbus article that described the incident. Shortly after touchdown, having deployed reverse thrusters, a pilot initiated a go-around but the aircraft veered to the left and did not climb as expected as only the right engine had engaged – the left engine remained idle. This caused the aircraft to veer to the left and the left wing only cleared the ground by 1ft before being able to land safely on another runway.

Dewi explained what happened around the logic of each engine's Electronic Control Unit (ECU), which exhibited different behaviour as each engine acts independently. Due to differences in each engine's sensor reading for 'wheels on ground', the left engine considering it was still in flight did not stow the reverse thruster, while the right engine sensing that both wheels were on the ground did allow the reverse thrusters to be stowed.




Dewi said this was example of the Byzantine General's problem with the wheel sensor being the 'traitor'. He noted that neither reports mention the Byzantine general solution but Airbus has since performed a software update.

Dewi said the paper is well known but not applied by certification authorities, perhaps due to the age of the paper, but he noted the number of citations are increasing especially in the last 10 years mainly coming from blockchain. Dewi said he is not claiming that systems need to be redesigned to Lamport's solution but it seems to have been forgotten that the problem exists. Dewi concluded by asking how we can raise awareness of this problem and suggested it could be taught at symposiums, in academia and could be explored during design review. However, he said it left him with a wider concern about what other problems may exist with solutions that we don't know about – the unknown unknowns – and what steps should be taken to uncover them.

## Safety in unmanned aviation – how much have we got and much do we need?



Steve Wright from Wright Airborne Computing introduced himself as previously working for Rolls-Royce, Airbus and Universities as a lecturer and said he has now moved into Uncrewed Aviation. He previously worked on large aircraft but is now working on uncrewed aircraft "crazy" vehicles.

Steve showed a video of 'drones through the internet' where drone data was being transmitted between France and the UK in real-time.

Steve also showed some of the other unusual projects he has been working on:

- Airborne robotics – a drone that unfolds itself and disguises itself as a brick
- Medical package delivery for the military
- Counterdrone aircraft

There have however been setbacks and Steve showed videos of drone crashes and battery fires and noted that the time imperative is strong and so failures are inevitable when 'pushing'.

Steve noted that reliability is never 100% especially during development and it can be fun building systems in 'wild west' environments, but there is a serious note and we are ultimately trying to preserve life. Steve covered differences between his old and new lives:

- **Old**: Meticulous design and tested, beautifully made, expensive and reliable
- **New**: Drone aero tech – multiple technologies that have evolved, borne out of mobile phone development. Software that should be £1M but is open source on the internet.



Steve then gave a live demonstration of a small toy drone and said it was an extraordinary piece of kit as it crams in all the usual flight systems you would expect in a conventional aircraft.

Steve then talked about aerospace reliability and showed that for conventional aerospace accidents the tolerable failure rate leading to catastrophic events is $10^{-9}$ per flight hour. This is compared with toy and commodity drones that may be as low as $10^{-1}$ for the one he had just flown. Steve illustrated the concept of the *gap of the six zeroes* showing a graph comparing reliability of vehicles from toy drone through to aircraft such as F16 and Predator drones and highlighted the gap between drones and conventional aircraft.

Steve made the distinction between availability – "does what we want" and integrity " doesn't do what we don't want" and noted the latter can be a counter-intuitive way of thinking. He said that the approaches to regulation are different between crewed and uncrewed aircraft with the latter focussing more on containment of accidents (via regulation CAP 722 and Open-A3-1), so the vehicle is expected to fail.

The environment is critical as well as aircraft performance and some companies like Manna are exploring these with urban deliveries on the west coast of Ireland – meeting the two considerations half way by boosting aircraft reliability and modifying the mission to reduce risk.

Steve then presented his personal opinions:

- It is not whether you crash but where and how. Availability is very low and integrity is more important
- "Flyaway" is a more important risk to avoid than crashing
- Safety personnel are disconnected from operations, so revert to no crash mindset.

Steve's final thoughts were:

- Crewed aviation: it's all about the vehicle
- Uncrewed: it's about mission plus environment
- Crewed/Uncrewed want same thing (want to operate safely) – but the debate is how we get there.

Steve finished his talk by describing his ground-breaking work with Aquark Technologies that involves dropping a scientific package in free fall (containing supercooled Rubidium atoms) to determine the performance of the apparatus in a zero-g environment.

## Things That Changed My World

At the end of the first day, Tom Anderson chaired an informal session with Phil Koopman, Yvonne Oakshott and Harold Thimbleby acting as panellists. Each panellist had three opportunities to tell stories that were challenging, important or entertaining and that influence their system-safety world. The audience then voted on their favourite story.

Round 1 was won by Phil (47% votes) who told a story about an 8-reactor nuclear aircraft carrier that he had worked on previously.

Yvonne won the second round (44% votes) with her moving story around the crash of an AW609 tilt rotor aircraft during development.

The final round was won by Phil again (72% of votes) with a story about the challenges he faced when he was tasked with investigating an Uncommanded Acceleration (UA) event by inspecting source code.

## Events

The morning of the second day started with a number of special events:

- Walking Tour of Bristol
- Bristol Cathedral Tour
- Bristol Museum Tour

These were all well-attended with all events being fully subscribed.



## System Safety for Complex Projects – The Crossrail Approach



The second day's keynote was given by Hayat Zerkani discussing the establishment of Crossrail and the building of the Elizabeth Line.

She said the line is mostly deployed in tunnels, with 10 new stations (9 subsurface) with portals and shafts for evacuation and emergencies, platform screen doors and step-free access throughout with a focus on a Central Operating Centre.

All this has been built by different joint ventures and needed to link with existing infrastructure.

The system is highly complex with many challenges:

- Noise reducing rails, non-traction power, UPS for critical systems
- Forced ventilation for cooling, control station and fire control
- Communications-based train control (CBTC)
- New trains
- European Train Control System (ETCS)
- Automatic Warning Systems (AWS)
- Signalling systems can change between different sections
- Rail milling trains to smooth tracks and other types of maintenance vehicle
- Regulations from Interoperability Directive 2008/57/EC, Safety Directive 204/49/EC and the Common Safey Method for risk ((EU) No 402/2013, CSM-RA)

To deal with this complexity, a more progressive handover was seen as more practicable rather than one as originally thought. Three stages were established:

- trial running
- trial operation
- entry into passenger service

Not all systems would need to be in place during non-passenger trial running.

Also a number of activities to ensure consistency were undertaken, such as developing a project-wide hazard record (PWHR), which was a key management tool to record hazards and safety justifications. This was implemented as a live database in DOORS[1] with a user friendly front end. This was maintained throughout the project and used as evidence for compliance with CSM-RA Regulations.

The safety justification had several parts: railway, contract and element level and it needed to bring these elements together including the station, shaft, portal and railway systems. It also essential to show integration with Network Rail so a Safety Justification was needed for the interface with railway companies and Crossrail.

This resulted in about 90 documents between contractors and Crossrail that needed to be maintained. Key success factors were:

- Progressive Assurance – deferring of some functionality to later configurations
- Safety Justification with dependencies, dependencies agreed with stakeholders and agreement of mitigations
- Ensuring consistency across all contracts was essential
- There were 600 dependencies so robust controls were required

Lessons that were learnt:

- reduce things into smaller more manageable components to help ensure consistency
- make use of tools
- focus on delivery
- be open with safety
- keep stakeholders engaged

[1] IBM® Engineering Requirements Management DOORS

## Panel Discussion "Can we make safe AI?"



The day concluded with Alan Simpson hosting a panel discussion with six experts in AI answering questions from the audience.

The following questions were covered:

Q1: Can we make AI safe?
Q2: How do we make AI safe?
Q3: How do we improve the competency of those developing AI systems?
Q4: What is the degree of Functional Safety in AI regulation?
Q5: Can you provide an example of the most complex AI system that has a safety case now?
Q6: Should individuals/the community be more proactive in highlighting the dangers of AI?
Q7: How do we prove AI systems are working correctly in law?
Q8: What might limit the arguably unbridled expansion of AI systems?
Q9: Would you be happy being driven by an Autonomous Vehicle?
Q10: Why do you trust Autonomous Vehicles less than high risk human drivers?
Q11: Can use of analysis techniques prove the absence of defects in a formal AI model?
Q12: What could ensure ethical use of ML and would that achieve safety as a consequence?
Q13: What advice would you give to the Safe AI Working Group?

You can see videos of the panel's responses to these on the event's Programme page on the SCSC website: https://scsc.uk/e1007prog

## Poster Session



The days talks were followed by a poster session in the main exhibition area where individuals and members of the SCSC Working Groups presented posters summarising their work on large presentation boards. There were many posters this year, which meant spreading the stands across several rooms. Some posters were also interactive inviting delegates to contribute their own ideas on the particular poster subject.



Posters were presented on the following topics:

- Dynamic Risk Assessment in Automated Vehicles: An elaborated Approach to HARA
- A Bow Tie Approach to Quantitative Risk Assessment and Control for Maritime Vessels
- Towards Trustworthy AI: Legislation, Techniques and Challenges
- SafeLLM: A Novel Framework for Safety Evaluation of Large Language Models: A Case Study of Offshore Wind Maintenance Planning
- Using Rust for Safety-Critical Systems
- Ensuring SOTIF for Off-Highway Autonomous Vehicles
- Adaptive Safety Measures: A Concept to Optimize Safety in Automated Driving Systems
- Safety and Complexity (Paper: How do we make the case for "Safe" software and AI systems? - the Horizon Example)
- IEC 63187: Systems engineering – System safety Complex systems and defence applications
- Safety and Complexity



And posters from the Working Groups:

- Safety Culture Working Group
- Assurance Case Working Group
- Data Safety Initiative Working Group
- Multi-Core Working Group
- Systems Approach to Safety of the Environment Working Group
- Safety of Autonomous Systems Working Group
- Service Assurance Working Group
- Safer Complex Systems Working Group
- Security Informed Safety Working Group
- Ontology Working Group

See page 51 for more details of most of the posters.

## Safety Timeline

A new feature for this symposium was the Safety Timeline where delegates were invited to add safety events that are important to them to a timeline stretching from 1400 to the current year. This was for example, filled by notable accidents such as Chernobyl (1984) and the Challenger Disaster (1986). The intention is to present the captured information in an interactive form on the SCSC website so that the timeline can be explored in more detail and new events can be added.




## Banquet

The end of the second day concluded with the traditional symposium banquet in the fine Palm Court Room at the hotel.

After an excellent meal, the attendees were entertained by Simon Davies from Vertical Aerospace who gave a talk and presentation on aviation and his role as test pilot for the new class of electric Vertical Take-Off and Landing (VTOL) aircraft currently under development by his company amongst others.













# The 'Golden Thread' of Information: A Systems Approach to Construction

The first keynote speech on the final day was given by Mikela Chatzimichailidou, Professor in Design for Mobility, Health, People and Society at UCL.




She covered three main areas in her talk:

- Background: motivation and literature
- Applied research: scope and publication
- Findings: Benefits and Challenges

**Background**

Mikela started by saying that construction came into her life after the 2017 Grenfell tower block fire that she witnessed firsthand from another building.

She said that the resulting independent review of building regulations and fire safety (the Hackitt Report), concluded that there were four main problems:

- Ignorance, regulations are not always read and often misunderstood and misinterpreted
- Indifference and commercial imperatives to deliver quickly and cheaply
- Lack of clarity on roles and responsibilities
- Inadequate oversight throughout the lifecycle

The proposed solution is to start thinking about the building as a system (it has a life) and she introduced the report's "golden thread of information", this being a tool – intended to be a digital platform – to manage all buildings as systems and to allow people to use digital records (acting as a database and single source of truth) to safely and effectively design, construct and operate residential buildings.

**Applied Research**

Mikela then went on to say that there is much innovation in modernising methods of construction, such as modular construction (constructing components offsite and assembling), sustainable building and 3D printed houses.

There is however debate as to whether Modern Method of Construction (MMC) building is safe and so Mikela along with an industrial research student undertook research (including a review of 74 relevant papers) to examine the link between Building Information Model/Modelling (BIM) systems, safety management and MMC. Mikela described BIM as a process with the output being the digital twin – the virtual representation of the system to be operated.

A BIM system can enable the duty holders to ensure accuracy and quality of design and construction, which are crucial for building-in safety up front and gives stakeholders access to a database from which to draw evidence.

Key findings from the research were categorised into three areas: training & warning, safety inspection and crane management.

**Training:** BIM can enhance training by supporting a virtual, interactive learning environment.

**Safety Inspection:** From manufacturing to construction, BIM-based dynamic automatic detection systems can effectively identify possible safety risks in modular construction.

**Crane Management:** operators and managers can more intuitively and accurately analyse potential conflict factors and find unreasonable lifting plans, thereby avoiding casualties caused by crane collision.

BIM therefore provides an opportunity to link worker safety and building safety under system safety to achieve the "golden thread of truth".

Mikela provided some conclusions:

- Buildings are complex
- Safety is a system's property, an emergent property
- Manage safety through requirements that are traceable throughout the building lifecycle
- BIM can be used to achieve the golden thread
- The golden thread brings together worker safety and building safety under system safety

There are challenges however, and Mikela provided eight challenges such as data protection, data assurance and skills shortage.

Mikela concluded with perspectives that are important not to forget:

- Digital tools are not here to replace human thought, critical thinking and evaluation still need to go into safety management; we should treat these as decision support tools
- Need to be confident about data assurance, its provenance, who has access and for how long and data quality
- Use of tools should be accompanied by contract commitments, for example, by putting use of BIM into contracts
- Seeing is believing – target and share successful case studies but be transparent
- It's a team effort, policy and practice need to go hand in hand
- One size doesn't fit all – be mindful of project specifics and nuances
- Need to remember the contribution being made in social/public values

## Safe by Design – Autonomous Vehicle Systems



Adiac Aguilar from Volvo Cars works on electronics and software platforms for Autonomous Vehicles (AV) and he gave his keynote talk on the 'Safety by Design' methodology. Adiac presented remotely having been unable to attend the symposium in person.

Adiac started by introducing the foundations of Safe by Design. This is based on a safety blueprint of the methodology's process and principles than involves safety culture, Continuous Improvement (CI) and human/machine/environment considerations.

The process starts with Hazard/Threat Analysis and covers System Safety and Cybersecurity streams.

Safety (Assurance & Risk) Management Systems are required to ensure a Safety Management System (SMS) is established covering Safety Policy, Safety Processes, Safety Risk Management, Safety Assurance and Safety Product Development.

**Architecture Design for Safety**

Adiac showed a high-level block diagram for an AV Architecture comprising:

- Communication
- Decision Making
- Perception
- Power Management and Control

He stressed the importance of designing robust and fault-tolerant architectures to mitigate safety risk.

**V&V for Safety Assurance**: Adiac also stressed the importance of rigorous testing and validation to verify the effectiveness and reliability of safety measures. Integration of V&V throughout the development lifecycle is also important.

**Safety Culture and Organisational Practices**

It is important to foster a strong safety culture to promote safe behaviours and attitudes. Integration of organisation practices and policies will support and reinforce safety culture. Adiac discussed the Safety Culture Framework covering:

Leadership Commitment, Employee Engagement, Communication, Safety Training, Continuous Improvement, Organisational Practices, Incident Reporting and Safety Committee.

**Ethical Considerations**

Adiac concluded by covering ethical considerations in three section:

- Ethical Dilemmas: trolly problem, allocation of responsibility between human and Artificial Intelligence (transition between level 3 and 4)
- Ethical Principles: safety transparency, accountability, fairness, privacy and social responsibility
- Ethical Frameworks: IEEE ethically aligned design framework, ISO/other standards, NHTSA Federal Automated Vehicle Policies

## Breaking the Tyranny of Net Risk Metrics for Automated Vehicle Safety

The last keynote speech[2] was from Phil Koopman from Carnegie Mellon University.




Phil noted that all Autonomous Vehicle (AV) manufacturers advertise as having safety as their first priority and the messaging that he is seeing is that these companies are saving lives, but the reality in his view is that a lot more miles will be required to prove it with sufficient confidence.

Robotaxi, for example, claimed in 2023 that humans are 'terrible drivers' citing 92% fewer collisions but the statistic only included 'at fault' accidents and hence, for example, did not count left turn faults so the statistics are potentially skewed.

There are however counterexamples where, for example, a Cruise robotaxi rear-ended an articulated bus seemingly ignoring the back half of the bus despite having LiDAR sensors. Phil said that "Robots won't drive drunk but will drive stupid"; humans will make same errors but for other reasons.

Further occurrences took place in August 2023 when an AV drove into cement and a crash into a fire engine resulted in the company being told to halve their fleet. The messaging that 'robots were good and humans terrible' has set companies up for a fall.

In San Francisco, Waymo have a better record but are not blame free. Phil noted that they are not required to report things where people were not injured but there have been safety-related incidents where incapacitated vehicles stop access to emergency vehicles.

[2] Phil's talk is based on a paper in the Safety Critical Systems eJournal, "Breaking the Tyranny of Net Risk Metrics for Automated Vehicle Safety" by Philip Koopman and William H. Widen https://scsc.uk/scsc-191

Nobody knows when AVs will be safer than humans but Phil believes companies are overstating the position and in fact possibly distorting the story. Phil estimates that 300+millions miles are required but they have only 5-8 at the moment.

A Cruise vehicle in Aug 2023 also drove into a road closed area and Phil asked if that was safe; there is more to it even when no harm is done.

The drive to demonstrate that AVs are safer than humans is called Positive Risk Balance but Phil said there is a lot more to it and there are other factors that need to be considered:

**#1 Positive Risk Balance (PRB):** means safer than a human driver, but which: (under influence, speeders, distracted, drowsy, older drivers, state and infrastructure) – national average are meaningless. What kind of vehicle – those with auto breaking are safer rather than older vehicles.
**#2 Avoid risk transfer**: are children more at risk, or cyclists? There is a problem if some specific demographics will lose out so we shouldn't increase a group's risk and so adds additional constraints on safety from PRB.
**#3 Avoid Negligent Driving:** establish a duty of care to other road users. There is no precedent for an AV. Product liability different from tort and liability. Should be similar to a human driver being negligent. Miles driven is no protection from culpability.
**#4 Standards and expectation on safety**: only one of the standards are currently required.
**#5 Fine-grain risk & regulators**: want to avoid regulatory recalls. Recalls always relate to specific behaviour such as rolling through stop signs; saving lives is not enough to avoid recalls. Regulators are struggling to predict safety outcomes in advance.
**#6 Ethical & Equity Concerns:** Ride Hail made promises with disappointing results. Labour issues, cheap taxi undermines safety, ethical & related concerns eg: testing risk imposed upon vulnerable people and no independent technical oversight is required.
**#7 Sustainable trust**: trust-degrading rhetoric. Unreasonable claims and relentless blame on humans. Lobbying for municipal pre-emption. Need more trust and need to talk with and not "at" with more transparency on incidents and corrective actions. Need to state release criteria and performance metrics.

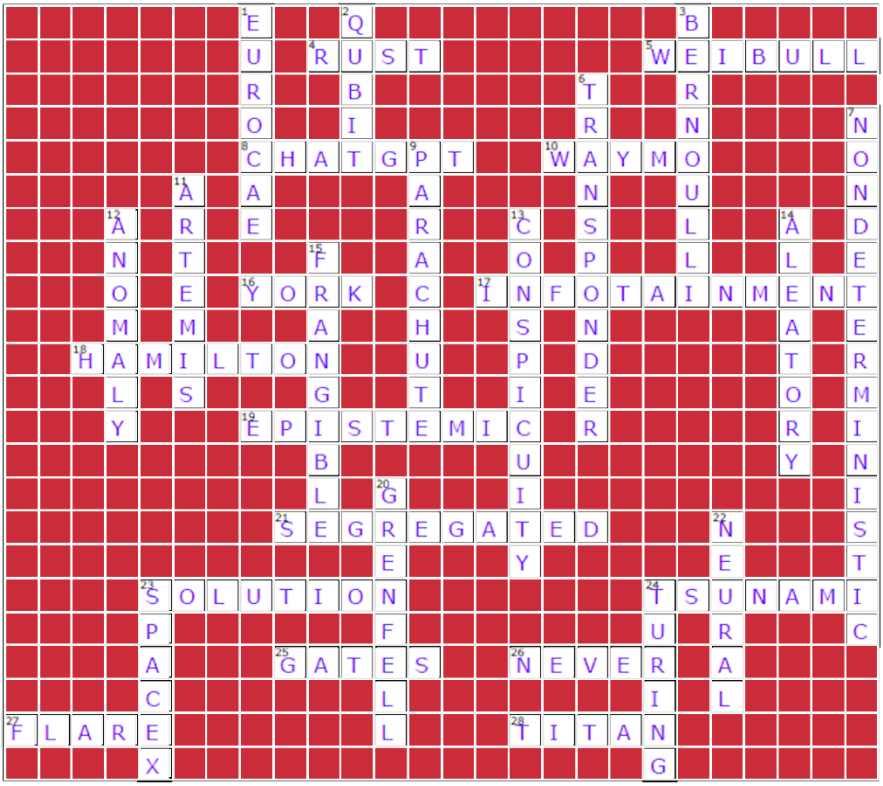
Phil concluded by saying that if you want to claim an AV system is safe then all these points need to be addressed.

# Crossword Competition




The SCSC Safety System crossword was first published in the Oct 2023 edition of the newsletter and was also given to delegates to complete during the symposium. There were 13 entries in total, and the winner was drawn from the correct entries in the 'hat' by Mitali Atkins with the winner receiving a £250 gift voucher.

The answers to the crossword are given below.



# Supporting the Development of an Inclusive Tech Industry

Women's Tech Hub grew from our shared love of working in tech, alongside our awareness of the lack of diversity. Our mission is to support the development of a tech industry that is inclusive of all so that it reflects the wider society.

Women's Tech Hub exists to encourage none-included groups, primarily gender-related who are in, or interested in tech to help support and develop their careers within the industry.

We aim to create a support system for women in tech, focusing on practical strategies such as workshops, career support, peer mentoring, informal discussions and signposting to training and other career solutions. We run regular workshops, meetups and events such as our supportive careers fairs and small conferences. We recognise that the more we support each other, the more we can get out there and inspire others.

## Networking



We support via our networking events as well as signposting other local tech-related groups which we often co-host such as 'CodeHub' and the 'Bristol Data and Machine Learning group'. We aim to help local tech companies create a more diverse and inclusive workplace culture that benefits everyone, through running the local chapter of Girl Geek Dinners and our careers fairs.



## Our events

We run regular events and networking opportunities including Workshop Wednesday and Girl Geek Dinners. We also try to keep an eye out on events of interest including tech conferences, meetups and any events that we run for our partners.

## Support



Women's Tech Hub provides a range of support services for women in tech, aimed at increasing diversity and gender equality in the tech industry.

## Recruitment



We realise that part of the issue with seeing more diversity within the tech industry is related to the recruitment pipeline. We have considered this and feel that the best option is to understand the pipeline issues and remove them as much as possible. For that we offer lightweight and reasonably priced jobs boards and passive recruitment. Rather than charging you an arm and a leg we simply act as match makers and advise on both sides of the recruitment process.

**www.wthub.org**

# SSS'24 Poster Sessions

As mentioned in the SSS'24 Symposium Report on page 35, the second day's talks concluded with a poster session where delegates were able to see a visual snapshot summarising the work and developments of particular groups such as the SCSC Working Groups. Each poster was attended by representatives from the group, who were happy to discuss the work with delegates in more detail.

Each poster is reproduced in the following pages for reference and a full pdf version of the poster can be seen by clicking on each image. The posters are:

- Dynamic Risk Assessment in Automated Vehicles: An elaborated Approach to HARA (Nikita Bhardwaj Haupt, Peter Liggesmeyer - RPTU Kaiserslautern)
- Towards Trustworthy AI: Legislation, Techniques and Challenges (George Mason and Greg Chance - Frazer-Nash Consultancy)
- SafeLLM: A Novel Framework for Safety Evaluation of Large Language Models: A Case Study of Offshore Wind Maintenance Planning (Connor Walker, Callum Rothon, Koorosh Aslansefat, Yiannis Papadopoulos, Nina Dethlef - AURA CDT, University of Hull)
- Using Rust for Safety-Critical Systems (Jonathan Pallant - Ferrous Systems UK Limited)
- Ensuring SOTIF for Off-Highway Autonomous Vehicles (Paul Darnell, Pavan Kumar, Ross McMurran, Ravindra Sabbella - 3SK)
- Adaptive Safety Measures: A Concept to Optimize Safety in Automated Driving Systems (Anil Ranjitbhai Patel, Peter Liggesmeyer - RPTU Kaiserslautern)
- IEC 63187: Systems engineering – System safety Complex systems and defence applications (James Inge)
- Safety and Complexity (David Slater, Cardiff University)
- SCSC Working Group developments: Safety Culture, Assurance Case, Data Safety Initiative, Multi-Core, Systems Approach to Safety of the Environment, Safety of Autonomous Systems, Service Assurance, Safer Complex Systems, Security Informed Safety and Ontology.

# Dynamic Risk Assessment in Automated Vehicles: An elaborated Approach to HARA

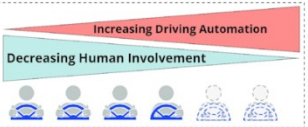
Nikita Bhardwaj Haupt[1], Peter Liggesmeyer[2]

Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

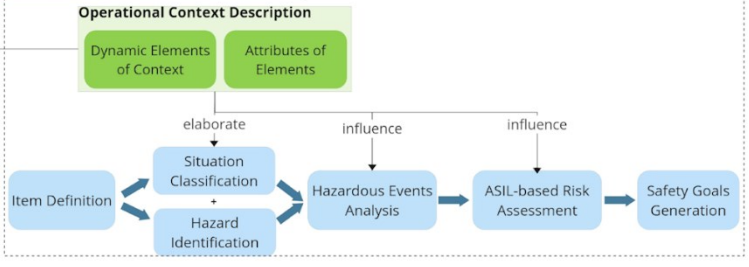
seda software engineering dependability

**Problem:** **Traditional HARA (as in ISO 26262)**

- relies on
  - worst-case assumptions about the situation
  - human driver fallback
- insufficient for higher levels of automation




**Solution:** **elaborated Hazard Analysis & Risk Assessment (elHARA)**




elaboration of operational context via

**Dynamic Element**
ENTITY_TYPE: pedestrian
ATTRIBUTE: • walkingspeed(wks)
• distancetoegoVeh(dst)
RANGE: wks ~ 0-3 mps
dst ~ 1-5 m

**Dynamic Element**
ENTITY_TYPE: vehicle
ATTRIBUTE: • drivingspeed(dvs)
• distancetoegoVeh(dst)
RANGE: dvs ~ 0-50 km/h
dst ~ 1-40 m

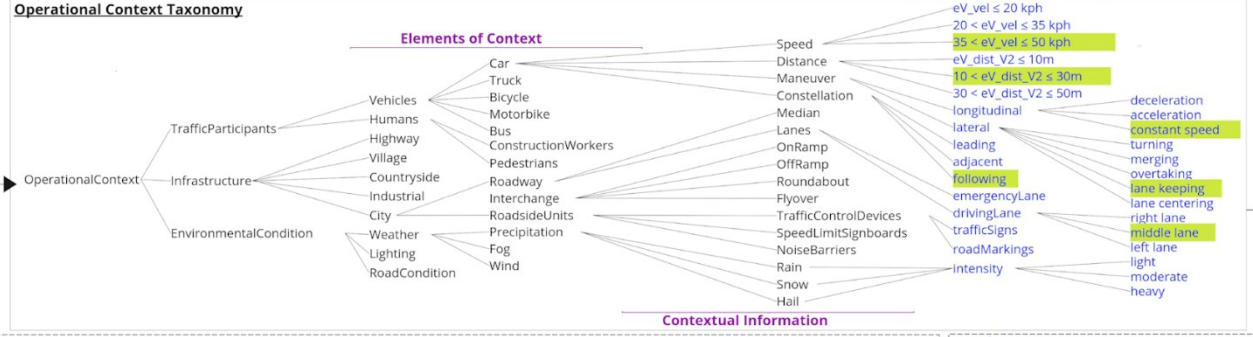
**Dynamic Element**
ENTITY_TYPE: stoplight
ATTRIBUTE: signallight (sgl)
RANGE: sgl: (red,yellow,green)

**Dynamic Element**
ENTITY_TYPE: vehicle
ATTRIBUTE: • drivingspeed(dvs)
• distancetofolVeh(dst)
RANGE: dvs ~ 0-50 km/h
dst ~ 1-40 m



generation of risk-relevant information

**Operational Context Taxonomy**




Elements of context, $E = \{e_1, e_2, \ldots, e_n\}$
For each element $e_i$ $A_{ei} = \{a_1, a_2, \ldots, am_i\}$ is the set of attributes for $e_i$
For each attribute $a_j$ of an element $e_i$ ,$C_{ij}^{ea} = \{c_1, c_2, \ldots, ck_{ij}\}$ is the set of contextual information, and
$I : E \times Ae_i \rightarrow C_{ij}^{ea}$ is an instance, a unique combination of a specific value for each attribute of each element.
→ Finally, the operational context, $OC = \{I_1, I_2, \ldots, I_p\}$, is a set of instances, where p = total number of instances.
→ In elHARA, Automotive Safety Integrity Level is estimated for each instance $I_m$, thereby allowing non-worst-cases about the context.

32nd Safety-Critical Systems Symposium

About authors:
Chair of Software Engineering,
RPTU Kaiserslautern, Germany
[1]nikita.haupt | [2]peter.liggesmeyer@cs.rptu.de

# Towards Trustworthy AI:
# Legislation, Techniques and Challenges

George Mason and Greg Chance

FRAZER-NASH CONSULTANCY
A KBR COMPANY

## Introduction

Most AI research has focussed on improving its abilities, i.e. the range of tasks it can do and how effectively it does them. By contrast, there is much less research into ensuring that AI is trustworthy, i.e. it is safe, fair, transparent, etc. Consequently, although AI is capable of many complex tasks – including high-risk tasks – there is limited assurance that it will do them without adverse effects.

This poster reviews the state of trustworthy AI in three areas:

1. Legislation: laws that jurisdictions are drafting for the responsible use and development of AI.
2. Techniques: methods to assess the risks of AI systems and ensure they are trustworthy.
3. Challenges: limitations of the legislation and techniques that must be overcome.

## Method

Data for this project was gathered in two stages:

**Stage 1:** Investigate AI legislation across jurisdictions worldwide. This includes AI-specific laws, general IT laws that can apply to AI, guidelines, and regulatory frameworks that are designed from the ground up to account for a wide range of AI uses and developments.

**Stage 2:** A review of research literature on techniques for trustworthy AI. This includes techniques to assess the risks of AI, such as subjective and objective metrics, and also techniques to ensure the principles of trustworthy AI are met.

## Results

### Legislation

Jurisdictions across the world are debating AI legislation with respect to mitigating its potential risks to society. Most legislation is in early stages of development; however, for some, ratification is imminent and others are already being enforced.

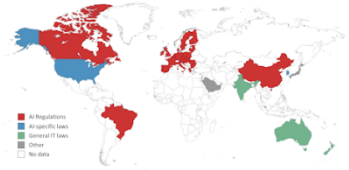



Figure 1: Jurisdictions where AI legislation is under development [1].

Generally, legislation fall into one of three categories:

1. Regulatory frameworks designed specifically for AI to apply to a wide range of uses and developments.
2. AI-specific laws targeted at particular uses and developments of AI.
3. General IT laws that can also be applied to AI.

Various guidelines have also been developed; however, these are voluntary – it is not mandatory to abide by them.

With some exceptions, most jurisdictions focus on several common principles to ensure trustworthy AI:

- Safety and accuracy: AI will behave as it is expected to.
- Transparency: how has AI made a decision.
- Fairness and equality: AI is free from racial, gender or other forms of discrimination.
- Accountability: who is responsible for the use of AI.
- Human oversight: decisions made by AI will align with human values.

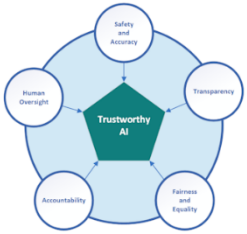



Figure 2: Common principles for Trustworthy AI.

### Techniques

Legislation prescribes *what* AI must or must not do, not *how* to achieve this. To satisfy legislation, techniques for risk assessment and to assure trust are also needed.

The level of risk of an AI-enabled system is assessed both subjectively, e.g. ethics, and objectively, e.g. functional.




Figure 3: Risk assessment spectrum, from subjective (left) to objective (right).

For more subjective assessments, such as ethics, human feedback is increasingly required, for example, in the form of questionnaires and polls [2].

For more objective assessments, such as functionality, various formal standards exist against which an AI system can be measured [3].

Different classes of techniques are required to ensure the principles of trustworthy AI are met. Possible techniques include:

**Safety and Accuracy:**
- Formal verification: mathematical analyses and proofs.
- Simulation: evaluate the system's behaviour in a simulator.
- Testing: check the system's code for implementation errors.

**Transparency:**
- Explainability: justify why AI has made its decisions.
- Traceability: examine all the factors that created the AI.

**Fairness and Equality:**
- Data sourcing: quality, representative data free from errors.
- Model selection: decisions are not biased, e.g. ethnicity.

**Accountability:**
- Redress: the ability to contest unjust decisions by AI.
- Auditing: algorithms, designs, data, etc. can be assessed.

**Human Oversight:**
- Human-in-the-loop: a human is required to enact decisions.
- Human-on-the-loop: a human can overrule a decision.

### Challenges

Some techniques for trustworthy AI are still in early stages of development: they may only offer vague assurances, may be too restrictive or have limited applicability [4].

Furthermore, even where effective techniques do exist, they may still be vulnerable to manipulation such that they provides assurances that are not true [5].

A further challenge is not stifling innovation of AI through excessive legislation [6]. Whilst it is important to ensure that AI is trustworthy, it is also important to not squander the untapped potential of AI because trustworthy AI legislation and techniques are too difficult or restrictive to incorporate.

## Discussion

As AI's ability to solve complex problems continues to improve towards human level (in some cases, already surpassing it), its potential to be applied to tasks traditionally done by humans is growing – this includes in safety-critical systems. However, measures to ensure that AI can be trusted in high-risk applications are not as mature; trustworthy AI is a relatively recent concept.

Legislation designed to ensure AI can be trusted is under development across the world at varying levels of progress. Although, the efficacy of the proposed legislation is yet unproven. In particular, it is unknown how well it will hold up against future forms of AI.

Techniques for AI trustworthiness is an active research area. Various metrics are being developed to assess the safety levels of AI and there has been a surge of attention into researching techniques to satisfy principles of trustworthiness. However, many trustworthy AI techniques developed so far suffer from being nebulous, vulnerable to attack or too restrictive. It may be some time before trustworthy techniques catch up to same level of maturity as AI's performative abilities.

A question becomes: how long would some actors be prepared to hold off on using powerful AI capabilities until the techniques necessary for them to be trustworthy are ready?

## References

1. IAPP Research and Insights (2023). Global AI Legislation Tracker. [online]. https://iapp.org/resources/article/global-ai-legislation-tracker/. Accessed 4th October 2023.
2. Chance, G., Abeywickrama, D. B., LeClair, B. et al. (2023). Assessing Trustworthiness of Autonomous Systems. arXiv:2305.03411v2 [cs.AI].
3. The Alan Turing Institute (2022). AI Standards Hub. [online]. https://aistandardshub.org/. Accessed 5th October 2023.
4. Hassija, V., Chamola, V., Mahapatra, A. et al. (2023). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. Cognitive Computation, pp. 1–30. doi: 10.1007/s12559-023-10179-8.
5. Noppel, M., Peter, L. and Wressnegger, C. (2022). Backdooring Explainable Machine Learning. arXiv:2204.09498v1 [cs.CR].
6. Department for Science, Innovation & Technology (2023). A pro-innovation approach to AI regulation. Great Britain. (CP 815).

## Acknowledgements

We thank all of our colleagues in the Digital Systems Assurance group at Frazer-Nash Consultancy for their valuable expertise and feedback throughout this project.

Particular thanks go to Rose Gambon, the group leader, and Lee Ramsay for their support that went beyond technical advice and leadership.

# Safety Monitoring for Large Language Models: A Case Study of Offshore Wind Maintenance

Connor Walker, Callum Rothon, Koorosh Aslansefat, Yiannis Papadopoulos, Nina Dethlefs

School of Computer Science, University of Hull - AURA CDT

## Abstract

It has been forecasted that a quarter of the world's energy usage will be supplied from Offshore Wind (OSW) by 2050 [1]. Given that up to one third of Levelised Cost of Energy (LCOE) arises from Operations & Maintenance (O&M), the motive for cost reduction is enormous. We recently introduced a specialised conversational agent trained to interpret alarm sequences from Supervisory Control and Data Acquisition (SCADA) and recommend comprehensible repair actions [3]. Building on recent advancements on Large Language Models (LLMs), we exapnd on this earlier work, fine tuning LLAMA [2], using available maintenance records from EDF Energy. An issue presented by LLMs is the risk of responses containing unsafe actions, or irrelevant hallucinated procedures. This paper proposes a novel framework for safety monitoring of OSW, combining previous work with additional safety layers. Generated responses of this agent are being filtered to prevent raw responses endangering personnel and the environment. The proposed framework has potential to utilise the O&M planning for OSW farms using state-of-the-art LLMS as well as equipping them with safety monitoring that can increase technology acceptance within the industry.

## Problem Definitions

**LLM Safety**
- Protection of maintenance personnel's exposure to risk, danger, or injury resulting directly from responses generated by the conversational agent.

**OSW Maintenance Planning**
- Ambiguous alarms, such as "WTG1A HV MAINTENANCE", presents difficulties in implementation of suggested actions and reduced safety critical awareness.

**Aims**
- Eliminate risks that exceed acceptable levels currently in the OSW industry.
- Filtering our conversational agents' responses to eliminate unsafe practices & maintain concise yet intelligible repair actions.

## Methodology

Figure 1 shows a block diagram of a process flow using a SafeLLM: Fine-tuned LLM within a safety-critical system, in the domain of OSW turbine OM. The process can be addressed in the following steps:

1. Input Prompt
2. Obtain LLM's Embeddings
3. Pre-defined Embeddings for Unsafe Concepts
4. SafeML* Score and Threshold Decision

*Safe Machine Learning (SafeML) uses statistical distance measures to estimate the ML Classifier accuracy.

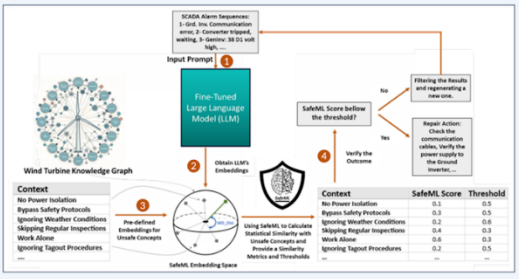



Figure 1. Diagram showing the overall procedure of SafeLLM

### Cosine Similarity

**Cosine Similarity** is defined as:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

Angle between sentence embeddings; Sentence embeddings; Vector norm of A and B

Visualised in similarity space:




- Sentence 1: "Complete maintenance on the wind turbine."
- Sentence 2: "Follow safe practices."
- Sentence 3: "Ignore safe practices."

Figure 2. Visualised Cosine Similarity

### Wasserstein Distance

**Wasserstein Distance** is simplified to:

$$W_1(P,Q) = \int_{-\infty}^{\infty} |P - Q|$$

Wasserstein Distance of p and q; Sentence embeddings; Absolute value of P − Q; CDF of p and q

Visualised in similarity space:

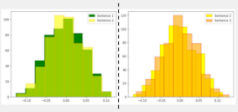


Figure 3. Visualised Wasserstein Distance

## Conversational Agent Example

Figure 4 suggests an example user interface for the O&M personnel. The safety filter proposed will prevent unsafe responses being shown to the end user, but would be alerted to the O&M manager to review.

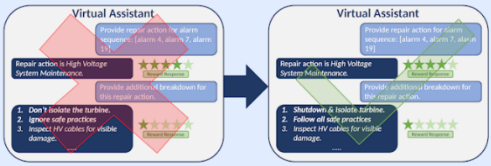



Figure 4. Example unsafe response being generated by the conversational agent.

## Results

Figure 5 below displays example confusion matrix heatmaps generated for 2 of the 10 categories. 20 sentences for each category are shown for simplicity.

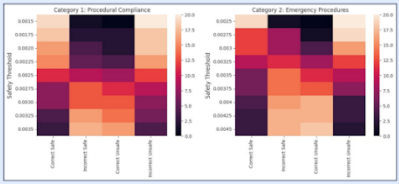



Figure 5. Sample Wasserstein Distance Confusion Matrix Heatmaps

Figure 6 below shows the false detection rate of Cosine Similarity and Wasserstein Distance for both unsafe and safe sentences combined at each safety threshold for all 10 categories.

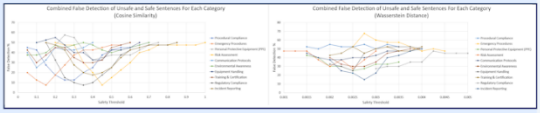



Figure 6. Categorised False Detection Rate for Cosine Similarity & Wasserstein Distance

Figure 7 presents the maximum overall accuracy of safe and unsafe sentences for both Cosine Similarity and Wasserstein Distance.

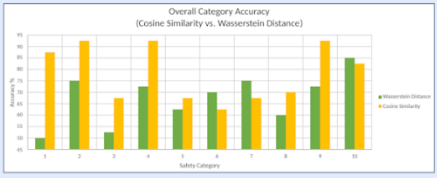



Figure 7. Overall category accuracy of Cosine Similarity and Wasserstein Distance

## Conclusion

The paper focused on defining Wasserstein Distance statistical distance measure as a framework to identify and eliminate unsafe concepts being suggested to maintenance personnel through our previously proposed Conversational Maintenance Agent. The aim of this work is to improve the safety of maintenance scheduling and compliance through decreasing reliance pressure on maintenance personnel. The results presented are preliminary, with scope for further optimisation in future. A foundation for using this framework has been developed with the results being comparable to current published work. Wasserstein Distance as a method is therefore believed to hold significant benefit in its application into the development of SafeLLM. As highlighted, a significant limitation to achieving validation of results is caused by lack of OSW specific data. It is hoped that through discussions and further development, we can validate the safety categories, as well as the unsafe concepts within each. This will allow for the creation of a dictionary containing concepts which align with current industry standards and processes.

## References

[1] Daniel Smith.
How offshore wind is evolving to hit 2050 net zero targets, 2023.

[2] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample.
Llama: Open and efficient foundation language models, 2023.

[3] C. Walker and et al.
A deep learning framework for wind turbine repair action prediction using alarm sequences and long short term memory algorithms.
In *Model-Based Safety and Assessment*, volume 13525, pages 189–203, 2022.

**Acknowledgements:** This work was supported by the **Secure and Safe Multi-Robot Systems (SESAME)** H2020 Project under Grant Agreement 101017258. We would like to thank EDF Energy R&D UK Centre, AURA Innovation Centre and University of Hull for their support.

ferrous systems

# Using Rust for Safety Critical Systems

ferrocene

**Ferrocene** is Rust, but **ISO26262 (ASIL D)** and **IEC61508 (SIL 4) qualified** by TÜV SÜD for use with safety-critical and functional safety systems. It is open-source and we offer long-term support on a commercial basis.

See https://ferrocene.dev for more details.

## Rust:

... is an open-source **systems programming language** designed to empower everyone to build reliable and efficient software:
**Safety - Performance - Productivity**

... is a **statically-typed compiled language**. Its powerful language-level static analysis prevents *use-after-free, double-free, shared mutability, threading race hazards*, and more. The libraries and binaries (known as crates) are composed of modules, which have clear type-checked public, semi-private and private APIs.

... is **cross-platform** and supports a wide range of Operating Systems:

- Windows, Linux, macOS
- FreeBSD/NetBSD/OpenBSD/Illumos
- VxWorks, QNX, LynxOS-178
- RTOS like ThreadX, FreeRTOS or RTEMS, and
- bare-metal platforms with no OS or heap

... includes:

- **cargo**, the package manager and build system,
- **rustdoc**, the documentation generator,
- **a test system,** which supports unit tests, integration tests and doc tests.
- A linter, **clippy**, which spots code which is technically correct but not idiomatic.

... uses LLVM and is a **cross-compiler** out-of-the-box. It supports a wide range of CPU architectures:

- Intel **i686** and **x86-64**
- **Arm** Cortex-M, Cortex-R and Cortex-A
- **RISC-V** 32-bit and 64-bit
- **SPARC**
- **PowerPC, MIPS**, and more!

```rust
//! Fuel monitor library.
//!
//! Provides the [`Monitor`] type to track fuel levels.
//!
//! Fuel levels are reported using the [`Reading`] type.

#![no_std]

/// Represents a single fuel reading
#[derive(Debug, Clone, Copy, PartialEq, Eq, PartialOrd, Ord)]
#[repr(C)]
pub struct Reading(pub u16);

/// Tracks fuel levels.
#[derive(Debug, Clone, Default)]
pub struct Monitor {
    /// A fixed-size buffer from the `heapless` project
    data: heapless::HistoryBuffer<Reading, 64>,
}

impl Monitor {
    /// Create a new fuel monitor, containing no readings.
    ///
    /// ```
    /// let fm = fuel::Monitor::new();
    /// ```
    pub fn new() -> Monitor {
        Default::default()
    }

    /// Add a fuel reading.
    ///
    /// ```
    /// # use fuel::{Monitor, Reading};
    /// let mut fm = Monitor::new();
    /// fm.add_reading(Reading(100));
    /// ```
    pub fn add_reading(&mut self, reading: Reading) {
        self.data.write(reading);
    }

    /// Get the mean fuel reading in the history buffer.
    ///
    /// ```
    /// use fuel::{Monitor, Reading};
    /// let mut fm = Monitor::new();
    /// assert_eq!(fm.mean_reading(), None); // No readings
    /// fm.add_reading(Reading(10));
    /// fm.add_reading(Reading(20));
    /// assert_eq!(fm.mean_reading(), Some(Reading(15)));
    /// ```
    pub fn mean_reading(&self) -> Option<Reading> {
        let sum: u64 = self.data.iter().map(|r| u64::from(r.0)).sum();
        let mean = sum.checked_div(self.data.len() as u64)?;
        Some(Reading(mean as u16))
    }
}

impl core::fmt::Display for Monitor {
    fn fmt(&self, f: &mut core::fmt::Formatter<'_>) -> core::fmt::Result {
        if let Some(mean) = self.mean_reading() {
            write!(f, "Fuel Monitor({} over {})", mean.0, self.data.len())
        } else {
            write!(f, "Fuel Monitor Empty")
        }
    }
}

/// C-compatible function to call [`Monitor::mean_reading`]
///
/// Returns `-1` if empty, or `-2` if NULL passed.
///
/// # Safety
/// You must pass a pointer to a valid, initialised, Monitor value.
#[no_mangle]
pub unsafe extern "C" fn fuel_monitor_mean(monitor: *const Monitor) -> i32 {
    // Check if pointer is not NULL
    let Some(monitor) = monitor.as_ref() else {
        return -2;
    };
    monitor.mean_reading().map(|r| i32::from(r.0)).unwrap_or(-1)
}
```

Rust macros derive the tedious boiler-plate code for you - without mistakes

These code blocks are automatically executed as part of the unit test suite. Your examples can never be out of date.

Rust has rich data formatting, without heap allocation.

Rust can use and export C functions

```
$ cargo test
   Compiling fuel v0.1.0 (/Users/jonathan/Documents/ferrous-systems/scsc-poster)
    Finished test [unoptimized + debuginfo] target(s) in 0.11s
     Running unittests src/lib.rs (target/debug/deps/fuel-5d9dd2cf32442f1c)

running 0 tests

test result: ok. 0 passed; 0 failed; 0 ignored; 0 measured; 0 filtered out; finished in 0.00s

   Doc-tests fuel

running 3 tests
test src/lib.rs - Monitor::add_reading (line 33) ... ok
test src/lib.rs - Monitor::mean_reading (line 44) ... ok
test src/lib.rs - Monitor::new (line 24) ... ok

test result: ok. 3 passed; 0 failed; 0 ignored; 0 measured; 0 filtered out; finished in 0.23s
```

# SOTIF for Non-Road Mobile Machinery (NRMM)

Paul Darnell, Pavan Kumar, Dr Ross McMurran, Ravindra Sabbella

## SOTIF

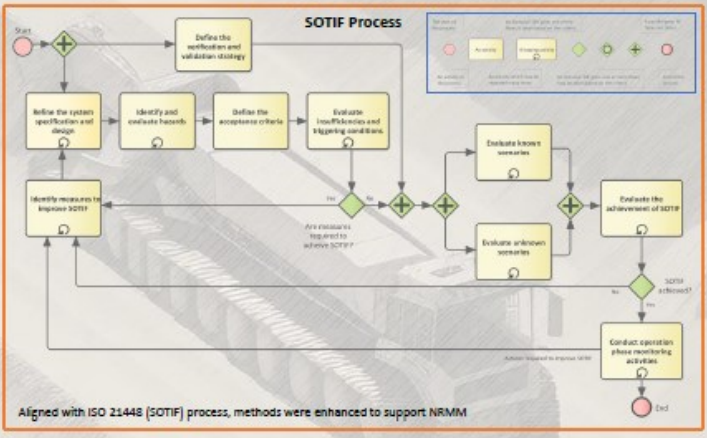
- Unlike traditional safety standards that primarily focus on addressing failures and malfunctions, **Safety Of The Intended Functionality (SOTIF)** emphasises assessing and mitigating risks associated with functional insufficiencies and user interactions to an acceptable level
- SOTIF standard (ISO 21448) is applicable to road vehicle systems which rely on sensing internal or external environment to build situational awareness

The goals of SOTIF




- 3SK team adapted and applied the SOTIF principles for NRMM

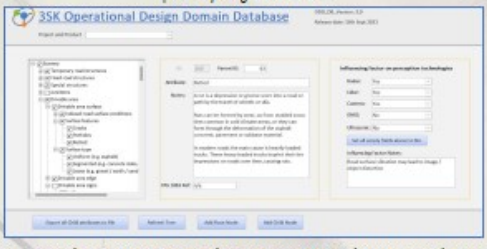
## SOTIF Process




Aligned with ISO 21448 (SOTIF) process, methods were enhanced to support NRMM

## Operational Design Domain (ODD)

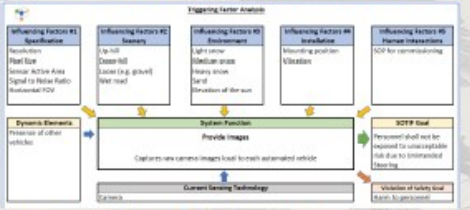
Operating conditions under which a given driving automation system or feature thereof is specifically designed to function




- 3SK's proprietary ODD database contains a superset of ODD attributes for all projects and products. This is continually enhanced with in-field learning
- Consistent with PAS 1883 (ODD taxonomy for ADAS), extended for NRMM
- ODD attributes are conservatively evaluated for their potential influence on perception technologies. Enables trade-off between preliminary system design complexity and ODD
- Be realistic on scoping the ODD. ODD has significant impact on design effort, V&V & cost!

## Analysis of potential insufficiencies and triggering conditions

Detect and address potential performance limitations, environmental factors, and foreseeable misuses that may result in hazardous behaviours identified in the HARA




- Established analysis methods, such as P-Diagram, FMEA, and FTA, are expanded and modified to systematically recognise and assess *triggering factors* and *insufficiencies*
- A unified analyses approach includes *functional insufficiencies* and *triggering condition*
- Insufficiency Management Strategy (IMS) include constraining ODD, updating standard operating procedures, upgrading perception systems, and exploratory testing
- Use ODD as starting point for identifying influencing factors of environment and scenery
- Constrain analysis to SOTIF relevant system elements and functions identified by FTA

## Acceptance Criteria and V&V

Approach is to break down the V&V task based on understanding of how SOTIF hazards may occur and lead to harm using the HARA & Triggering Factor Analysis

$$A_H = R_{HB} \times P_{E|HB} \times P_{C|E} \times P_{S|C}$$

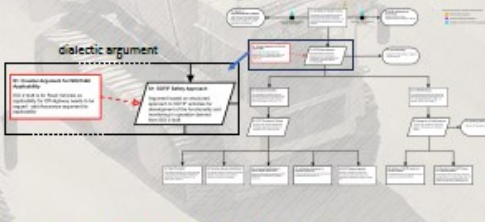
$A_H$ - Acceptable occurrence of Harm
$R_{HB}$ - Rate of occurrence of hazardous behaviour
$P_{E|HB}$ - Conditional probability of being in circumstances that can lead to harm if hazardous behaviour occurs
$P_{C|E}$ - Conditional probability that hazardous behaviour is not controllable in circumstances that can lead to harm
$P_{S|C}$ - Conditional probability of accidents of a particular severity will occur in these circumstances




- The V&V results should provide evidence for the calculation of the known risk
- AND the rate of finding new hazardous scenarios and their probability of causing harm to give sufficient confidence in the margin from the known risk to the acceptable risk
- This must include sufficient exposure to real world operation
- This approach leads to simpler tests focused on specific aspects of the system, which are applicable to many hazards and hazardous scenarios.

## Safety Case

Used to make the underlying argument of ISO 21448 (SOTIF) explicit and to ensure work-products were developed to provide evidence of compliance




- Concurrent Development of Product and Safety Case
- Use of Goal Structuring Notation (GSN) v3 including "Assurance Claim Points" to branch off such separate arguments from the main risk argument and "dialectic arguments" for review
- Use of Argument type indicators (not part of GSN standard)

Paul Darnell, Technical Director, paul@3sk.co.uk
Pavan Kumar, Delivery Head, pavan@3sk.co.uk
Dr Ross McMurran, Technical Specialist, ross@3sk.co.uk
Ravindra Sabbella, Practice Head, ravindra@3sk.co.uk

© 3S Knowledge Limited 2024

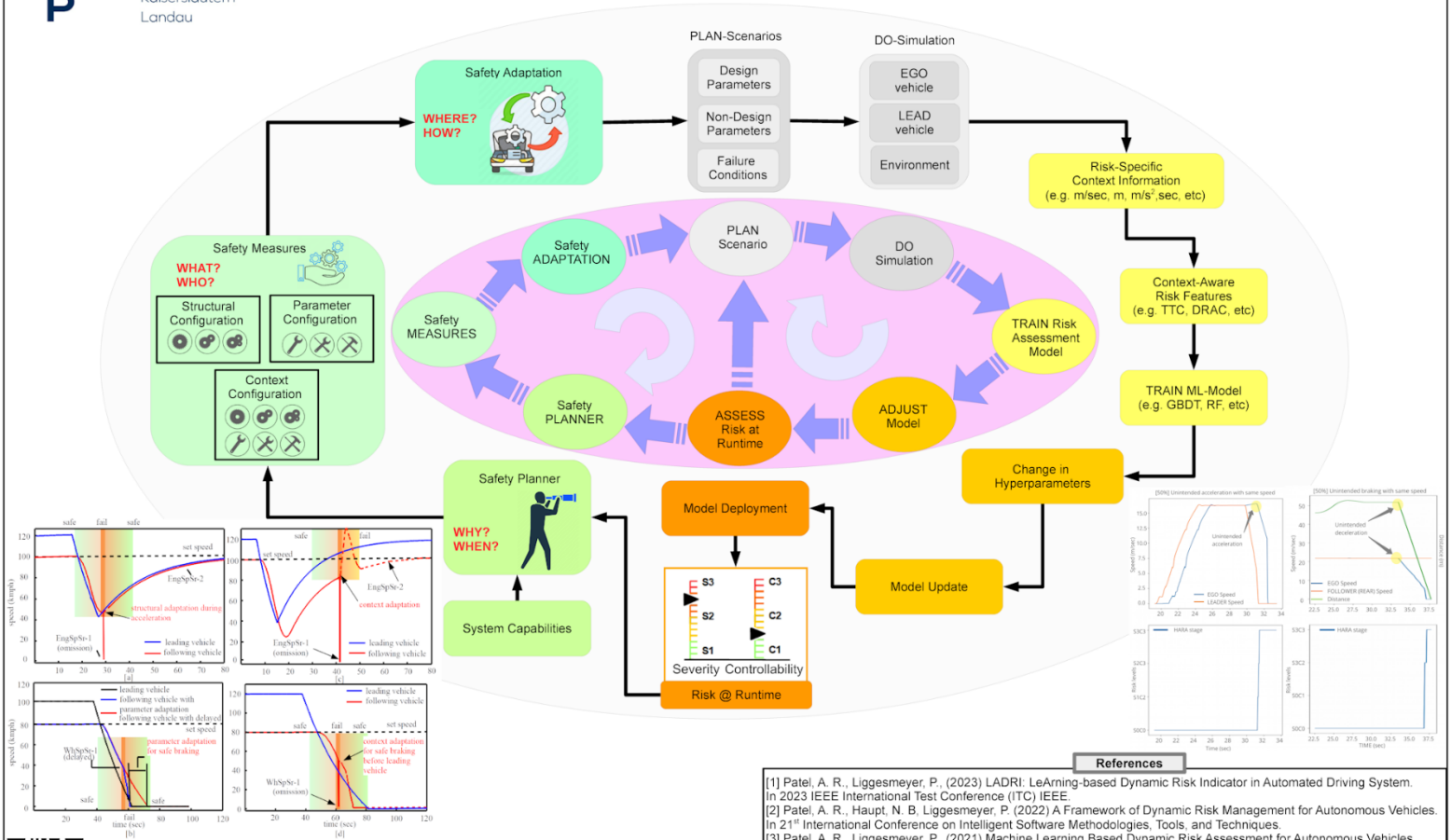
Innovation Centre, Warwick Technology Park
Gallows Hill, Warwick, UK, CV34 6UW
www.enquiries.co.uk

# Adaptive Safety Measures:
# A Concept to Optimize Safety in Automated Driving Systems

Anil Ranjitbhai Patel[1], Peter Liggesmeyer[2]

Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

seda software engineering dependability




**References**

[1] Patel, A. R., Liggesmeyer, P., (2023) LADRI: LeArning-based Dynamic Risk Indicator in Automated Driving System. In 2023 IEEE International Test Conference (ITC) IEEE.

[2] Patel, A. R., Haupt, N. B, Liggesmeyer, P. (2022) A Framework of Dynamic Risk Management for Autonomous Vehicles. In 21st International Conference on Intelligent Software Methodologies, Tools, and Techniques.

[3] Patel, A. R., Liggesmeyer, P., (2021) Machine Learning Based Dynamic Risk Assessment for Autonomous Vehicles In International Symposium on Computer Science and Intelligent Controls (ISCSIC)

[4] Patel, A. R., Haupt, N. B, Liggesmeyer, P. (2021). Prediction of Dynamic Adaptation Technique for Autonomous Vehicles using Decision Trees. In 29th Safety-Critical Systems Symposium (SSS'21). SCSC

[5] Patel, A. R., Haupt, N. B, Liggesmeyer, P. (2020) Exploiting Adaptation Behavior of an Autonomous Vehicle to Achieve Fail-Safe Reconfiguration. In 6th International Commercial Vehicle Technology Symposium Kaiserslautern

32nd Safety-Critical Systems Symposium,

Anil Ranjitbhai Patel[1], Peter Liggesmeyer[2]

[1]Participating Investigator and [2] Scientific advisor

Chair of Software Engineering: Dependability,

Rheinland Pfälzische Technische Universität Kaiserslautern, Germany

# IEC 63187

## Systems engineering — System safety Complex systems and defence applications

"A specialised instantiation of ISO/IEC/IEEE 15288 for the safety aspects of the lifecycle of systems, replacing IEC 61508-1 for the scope of complex systems"




IEC 63187-1 is currently at the Committee Draft stage of development, with publication as an international standard planned for late 2025, subject to agreement.

IEC TR 63187-2 is being drafted as a Technical Report to support implementation of IEC 63187-1.

### Motivation for a new standard

The market is evolving towards:

- More **complex** systems with **complex** functions and **complex** architectures
- **Fewer humans in the loop** to handle safety
- **Dynamically evolving risks**

Existing safety standards do not:

- Align with **system engineering**
- Address **multi-layered systems** recursively
- Capture **emergent properties** at system level (without failure)

### Objectives

Safety:

- Propose a comprehensive approach for **adapting safety requirements to risk**
- Propose an approach for **risk control** across the layers of a system
- Propose an approach to control situations of **dynamic risk**
- Maintain an open approach towards activities and assurance **outcomes**
- **Account for traditional approaches**, in particular quantitative, for **realised system elements**

System engineering:

- Propose an approach of **embedding safety engineering** in systems engineering (ISO/IEC/IEEE 15288)
- Propose an approach enabling **instantiation between system layers** and between supplier tiers **without limitations**
- Distinguish the system conceptual activities from the realisation of the system physical and logical elements
- Propose a seamless **interface with existing safety standards** for realising physical and logical system elements

### Fundamental principles

- **Risk-based** (ISO 31000): focusing on hazardous events and their consequences
- **Systems-based** (ISO/IEC/IEEE 15288): recursively applying systems engineering principles to complex defence systems
- Systems and **control theory**: considering interacting system elements
- **Integrating safety into systems engineering** as a fundamental part of engineering the system, with the flexibility to trade requirements
- **Appropriateness of approach:** addressing threats to safety from sources with different types of characteristic
- **Supply chain considerations**: addressed to apply proportionately over organisational boundaries
- **Solution independent**: regardless of the origin or lifecycle stage of realised system elements
- **Goal-based** providing a framework for achievement of safety objectives through a principled approach to engineering activities

### Procurement viewpoint

- Defence applications are subject to dynamic risks: detriments/losses, safety objectives and compromises depend on the operational context (CONOPS)
- Risk acceptability can only be determined on a case-by-case basis, by the organisation purchasing the system
- Definition of the detriments (losses) from the procurer's own viewpoint (i.e. "what is important")
- Convergence and control of Measures of Importance (MoI) through schemes agreed between parties (i.e. "how important is that?")
- IEC 63187 is entirely applied by each organisational Tier, giving a consistent approach
- Each Tier defines its own suitable life cycle on the basis of the 3 generic ISO/IEC/IEEE 12207 life cycles
- Tiers agree on the organisation of the interfaces, allowing consistent and traceable engineering



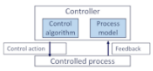
### Systems engineering viewpoint

**Control theory:**

- All loss domains (mission, safety, security, …) can be considered in decision making
- Centred on the Perception–Interpretation–Decision–Action loop.
- Safety viewed as the robustness of control under internal and external perturbations
- Modelling control structures allows scenarios that lead to undesired system states to be identified
- Humans are integrated in the control structures

IEC 63187 considers all the system life cycle processes from a safety viewpoint:

- Implementing a "Safety View" (as an aspect of speciality engineering, as per ISO/IEC/IEEE 15288:2015 annex E4)
- Supplementing each ISO/IEC/IEEE 15288 process with specialised requirements and informative notes, to clarify the outcomes needed from the safety viewpoint



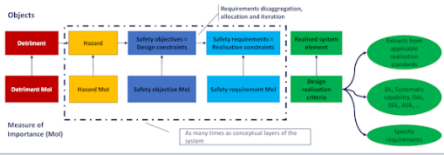

### Implementation viewpoint

- The actual realisation of the system elements produced is not part of the scope of IEC 63187
- Inputs will be defined based on detriments; safety objectives and the design constraints will integrate the elements necessary for the realisation
- The standard is limited to providing routes to specify and accept system elements, without entering into the details of their realisation or interfering with prescriptive standards for the production of system elements (e.g. DO178C, IEC 61508-3, IEC 61508-2)

| | Visibility of design and configuration | No visibility of design and configuration |
|---|---|---|
| Controllability of the design | Route 1 | Route 3 |
| No controllability of the design | Route 2 | Route 4 |

### Safety engineering viewpoint

- The IEC 63187 Measure of Importance (MoI) concept allows classifying the importance of various objects according to associated criteria/parameters and moderation factors
- Defining a MoI makes it possible to avoid the saturation of integrity levels that can occur when allocating requirements mechanically
- The IEC 63187 approach remains compatible with the fundamental principles of IEC 61508 and MIL-STD-882E




### Takeaway points

**Systems engineering**
- No mandated safety deliverables; safety outcomes are embedded in the systems engineering outcomes, open life cycle adaptable by each stakeholder organisation.

**Hazards, risks and detriments**
- Principles of control theory.
- Based on a unique concept to express objectives from all specialities and allow arbitration when necessary.

**Safety objectives and safety requirements**
- Dissociation of constraints on the system of interest (objectives) from the solutions satisfying them (requirements) and allows identification of emerging aspects.

**Measures of Importance**
- No predefined index (no equivalent to SIL, DAL, ASIL, etc.)
- Definition of normative requirements to allow stakeholders to define ad hoc Measures of Importance in a consistent global framework.

**Safety performance**
- Accounts for the fact that the system safety performance, if expressed only quantitatively as the sum of the realised system element failures, cannot represent the overall system safety.

**Safety assurance**
- Accounts for the fundamental difference between the necessary means to achieve safety objectives and the necessary means to guarantee their effectiveness.

### IEC 63187 development ecosystem

**26** **Plenary meetings** since 2018

**178** **National comments** received on draft







IEC 63187 is being developed by Working Group 18 of the IEC's TC65/SC65A technical subcommittee: Industrial-process measurement, control and automation – Systems Aspects. Members of the international working group are experts nominated by the national standards organisation of each country that takes part.

In the UK, working group members are nominated by the British Standards Institute's GEL/65/1 committee.



For more information, contact james.inge@scsc.uk

# Safety and Complexity

"If you can't model it, you can't manage it"

David Slater

## How do we make the case for "Safe" software and AI systems? - the Horizon Example

### Two Basic approaches

### 2. – System thinking



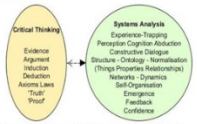


"A system is never the sum of its parts; it's the product of their interactions. -

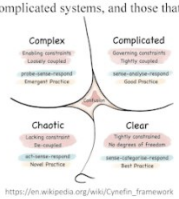
*Ackeff*

*Deming*

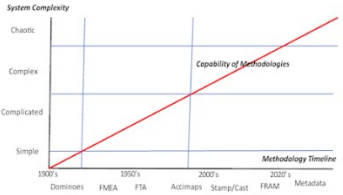
Management of a system requires knowledge of the interrelationships between all of the components within the system and of everybody that works in it.

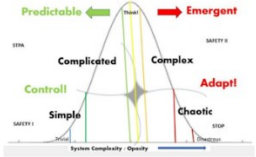


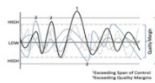

***Snowden in his CYNEFIN concept,*** recognised that we needed not just system thinking, but to recognise that there are fundamentally different behaviours in "predictable," but complicated systems, and those that are truly COMPLEX and can behave unpredictably.




https://en.wikipedia.org/wiki/Cynefin_framework







"NORMAL" Accidents

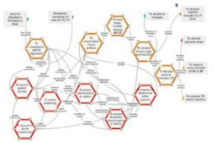
Complex

**Hollnagel** - visualised complexity as a system of interconnected, interdependent, and mutually interactive Functions

Accidents are the result of the interactions of systemic functions and performance variability; and can emerge unexpectedly when conditions "resonate". Also identifies Work as Done vs theoretical predetermined procedures, (WAD vs WAI).

SAFETY by ensuring success!



With the addition of assigning individual functional metadata, Hollnagel's FRAM methodology now enables tracking and calculating the interactions that stochastically affect systems behaviour





Complicated

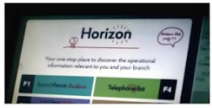
**Leveson** - Safety is a system control problem – extended Accimaps into the STAMP CAST map of management control loops

SAFETY by ensuring system is under control

### Suitable Methodologies for Accident / System Investigations

| Cynefin Category | Simple | Complicated | Complex | Chaotic |
|---|---|---|---|---|
| System / Accident Analysis Methodology | Heinrich Dominoes | Failure Mode and Effect Analysis- FMEA (Component ) | Swiss Cheese Model | Avoid |
| | Root Cause Analysis (RCA) | Fault Tree / Event Tree Analysis (FTA) | Failure Modes and Effects Analysis – FMEA (systemic) | |
| | | Bow Ties and Barriers | Risk Matrices | |
| | | Layers of Protection Analysis – LOPA | System Integrity Levels – SILs | |
| | | Probabilistic Risk Assessment- PRA | Bayesian Belief Nets | |
| | | STAMP/ CAST | Accimaps | |
| | | | SADT | |
| | | | FRAM | |

Increasing sophistication

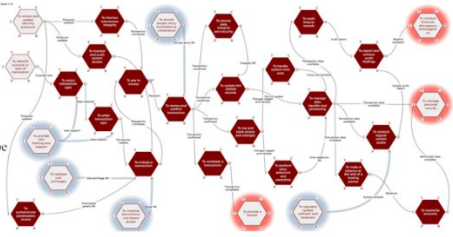



*"Programming is a human task and programmers make mistakes; an error rate in writing software code of 10 errors per thousand lines of code is considered good, 1 error per thousand lines is rarely if ever achieved." (Harold Thimbleby et al)*

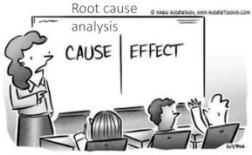
This Poster shows an attempt to address systematically the issues which are proving difficult in analysing and assuring the performance of safety critical software systems.

We made an initial attempt to develop and validate a better way to model and assure the performance of modern software packages by applying it to the Horizon software. It thus looks at the potential for applying more advanced methods of modelling and analysing these systems.

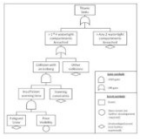
- The first approach investigated was the use of the Functional Resonance Analysis Method to build the system visualisations – models.
- Secondly since there have been concerns about how difficult it is to build the initial FRAM models which can then be used to examine in detail the behaviours possible in these complex systems.
- So the feasibility of using LLM's to produce initial outline systems models was explored.
- The results have been reported in a note to the Safety Critical Systems working group



### Root cause analysis



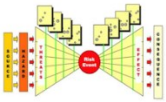

"Let's just stick with this for now. We can talk about unintended consequences later."



Logic trees



Heinrich's Dominoes



Bow ties and Barriers

Decomposition, Root cause analysis, Fault Tree Analysis, universal "laws"

SAFETY by technology, finding and fixing faults and failures - engineering out hazards

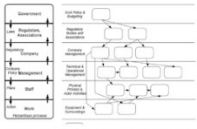
### Fixing "Human Error"

Humans as components,
– substitute their "failure rates" in Fault trees, etc,
– HTA, HFACS, CREAM, SHERPA

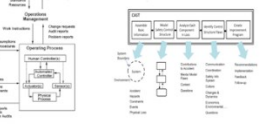
SAFETY by "Human Factors"
SOP's, SMS's, Training, Ergonomics.



Accidents are systemic "failures"
– **Perrow**, Complexity plus "tight" coupling in system interactions makes accidents inevitable.
Needs better "Management"?
Rasmussen Accimaps identify management "chains" and influencing hierarchies







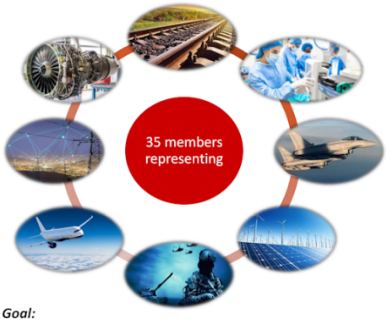
Poster sponsored by Ebeni

# Safety Culture Working Group

*"The safety culture of an organisation is the product of individual and group values, attitudes, perceptions, competencies, and patterns of behaviour that determine the commitment to, and the style and proficiency of, an organisation's safety management system (within the overall business management system)."*

*Adapted from ACSNI Human Factors Study Group: Third report - Organising for safety HSE Books 1993*

## Who we are




***The SCWG Goal:***

*'**Improve safety culture** in safety critical organisations focussed on product and functional safety, by **sharing examples and latest approaches** collated from real life case studies'*

## Our successes - 2023

### Four Webinars

- "Measuring safety culture: Practices and Pitfalls"
- "UK rail organisations approach the improvement of safety culture"
- "Health and wellbeing management of train drivers"
- "Safety culture in healthcare"



### "A position paper for assessing and managing safety culture"

- Providing guidance on the assessment and improvement of safety culture
- Summarises lessons learned on how to improve safety culture
- Demonstrates that safety culture influences all aspects of safety performance throughout the system lifecycle



## What does safety culture look like?

A positive safety culture is critical for successful safety performance across design, build, transition, assurance and operations. Here we explore what 'good' and 'poor' safety cultures look like

### Common elements of a 'good ' culture

Learning Culture → Leaders who demonstrate positive safety behaviours → Shared values, beliefs and norms → Openness, fairness and trust → Continuous improvement → Open and honest comms → Clear lines of accountability and responsibility → Questioning and adaptable culture

### Poor Culture Incident Case studies

**Poor safety culture within assurance:** a fire and crash of the RAF Nimrod Maritime reconnaissance/strike aircraft in 2006

The Inquiry reported that there had been **a failure of leadership** in producing the safety case, almost considered a 'tick box exercise'. Several previous incidents and **warnings signs were ignored**, risks not adequately managed and mitigated and a latent flaw was not adequately addressed

**Poor safety culture within Operations:** Mortality rates at Mid Staffordshire NHS Foundation Trust 2005-2008

The inquiry reported that '*conditions of appalling care were able to flourish*' and a lack of basic care of patients to persist was primarily caused by a serious failure on the part of the provider Trust Board.....and its failure to tackle an **insidious negative culture** involving **tolerance of poor standards** and **disengagement** from managerial and leadership responsibilities




## The focus for 2024

- Four new webinars for 2024
- The SCWG will be developing a new **Position Paper** which considers:

***A desktop study throughout key industries to identify and understand any trends between safety culture performance and events.***

## Interested in collaborating?

- We are always looking for new members to join our community
- For information on how to join the SCWG, visit the website:
- Or contact the group leader Michael Wright at michael.wright@scsc.uk

# Assurance Case Working Group

Find out more and contact us @
ACWG Website: https://scsc.uk/gc and GSN microsite https://scsc.uk/gsn

SCSC
SAFETY CRITICAL SYSTEMS CLUB
FOR EVERYONE WORKING IN SYSTEM SAFETY

The Assurance Cases Working Group (ACWG) was established in 2017 to provide guidance on all aspects of assurance cases including construction, review and maintenance.

It is intended that this be broader than safety, and address interaction and conflict between related topics. It addresses aspects such as Proportionality; Rationale behind the guidance; Focus on risk, confidence and conformance; Role of challenge, counter-argument and counter-evidence; and Treatment of potential bias in arguments.

The ACWG has established the GSN_SWG, working under its remit, to maintain the Goal Structuring Notation (GSN) Community standard.

***WG Membership:***

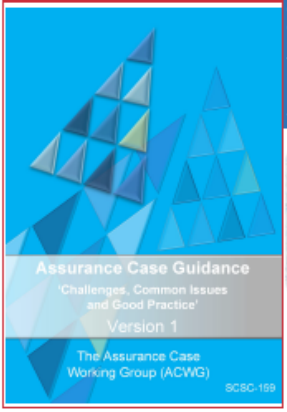
- *Active: 42*
- *Observers: 13*
- *Industries represented:*
  *Academia, Aerospace, ATM, Automotive, Consultancy, Healthcare, Maritime, Nuclear, PetroChemical, Rail*
- *Countries represented:*
  *Australia, Belgium, Canada, Germany, Japan, Norway, Poland, UK, US*

**Current Activities:**

The ACWG is working on guidance material for **Proportionality** and **Assurance Case through the Supply Chain.**

The ACWG is also forming plans to take the next steps for the AC guidance and GSN Standard. These plans are embryonic, and contributions are welcome.

## Publication




Guidance has been created to support the production and maintenance of assurance cases that are complete, coherent and proportionate to the confidence required. It is aimed at those who create, review, approve and use assurance cases and is not sector specific.
The guidance takes the form of a framework that provides a general introduction to the scope and context of assurance cases supported by papers that each address a specific guidance topic. Supporting terminology and references are provided.




***WG output:***

- *Assurance Case Guidance:*
  - *Published: Aug 2021*
  - *Downloads: 2986 (7 Dec 22)*
  - *Amazon Sales: 16 (16 Nov 22)*
  - *WIP: Planning stage for V2*
- *GSN Standard*
  - *Published: May 2021*
  - *Downloads: 4149 (7 Dec 22)*
  - *Amazon Sales: 12 (16 Nov 22)*
  - *WIP: Planning stage for V4*
- *Published documents are made available:*
  - *Under Creative Commons 4 license*
  - *In soft copy (A4-pdf) for **free** via SCSC website*
  - *In hardcopy (8.5"x11") via Amazon (search SCSC in books)*
- Online Resource (GSN microsite)

**Guidance topics in the current publication:**

- **Avoiding Bias in Assurance Cases:** identifies a number of common biases to be aware of (and thus avoid) when constructing, reviewing and using an assurance case.
- **Risk versus Benefit:** outlines the need to balance risk against benefit and how to tell when such an argument may be required.
- **Modular Assurance:** deals with the structure and presentation of large and complex assurance arguments to support reader comprehension and managing the impact of change.
- **The Risk-Confidence-Conformance Approach:** introduces the concept for clear presentation of arguments structured around three central themes.
- **Dialectic Arguments in Assurance Cases:** introduces the concept of dialectic arguments to add confidence to an assurance argument while it is being authored or during review/evaluation.

# Goal Structuring Notation (GSN) Standard

GSN is a graphical argument notation that can be used to document explicitly the individual elements of any argument (claims, evidence and contextual information) and the relationships that exist between these elements and the context that is defined for the argument.

The GSN Standard includes:

- Introduction And Concepts;
- Definition Of The Goal Structuring Notation;
- Guidance On The Development And
- Evaluation Of Goal Structures.

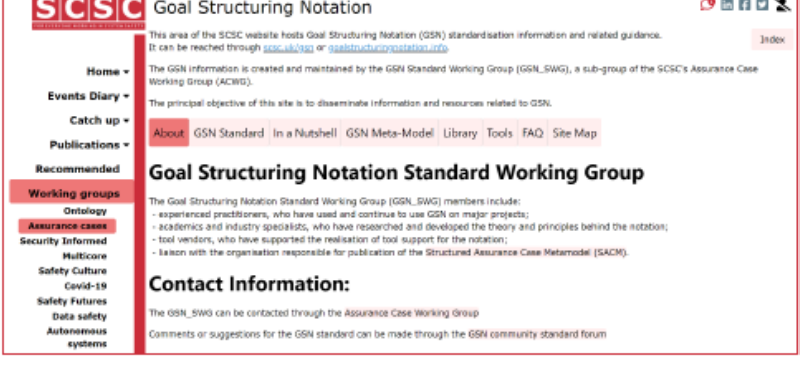
A complementary web-based repository of GSN resources (GSN microsite) has been created.

A meta-model of GSN has been developed, showing the relationship to Object Management Group's (OMG) Structured Assurance Case Metamodel (SACM).




Printing sponsored by:
Engineer for Safety and The Furious Engineer Limited
Independent System Safety Consultancy and Training

## Goal Structuring Notation Microsite https://scsc.uk/esn (also http://goalstructuringnotation.info)




## Publication



# Data Safety Initiative Working Group (DSIWG)

**Started: January 2013** **82 Meetings** **89 Members**



**Mission Statement:**
**"To have clear guidance on how data (as distinct from the software and hardware) should be managed in a safety related context, which will reflect emerging best practice."**

www.scsc.uk/gd mike.parsons@scsc.uk

## Recent Data Safety related accidents

### 2023 - NATS flight plan outage

A failure of one of NATS' flight planning systems to cope with a particular combination of duplicate waypoint data being fed into the system as part of one valid flight plan, caused both the primary system to fail and its backup to enter a fail-safe mode. The entire shut down process took seconds from the point of the flight plan being filled to entering fail-safe mode but recovery took longer due to difficulties identifying the problem. This included finding the specific data, isolating it and removing it in a controlled way, and then testing it to ensure it could be returned safely into operation. Thousands of passengers faced disruption as a result of the failure, with over 1500 flights cancelled and hundreds more delayed.

### 2023 - Post office Horizon scandal

In 1999, the Post Office began automating accounting processes at about 14,000 Post Office branches. This saw the introduction of a centralised computer system from supplier Fujitsu. This system replaced traditional paper-based accounting practices. But problems ensued, and there was a sudden increase in the number of subpostmasters suffering unexplained accounting shortfalls. Rather than investigate the problems and fix them, the Post Office blamed the branch operators, many of whom it prosecuted for financial crimes, with many more made bankrupt and sacked.

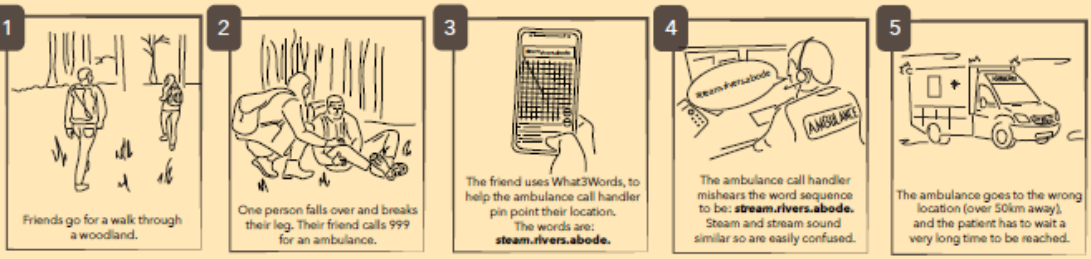



**"Data: The Lifeblood Of AI.**
**At the core of every AI system lies a fundamental truth: The quality and quantity of data it ingests are paramount to its effectiveness. In essence, data is the lifeblood that fuels AI algorithms, allowing them to learn, adapt and make decisions."**

**Rohit Sehgal, Founder and CEO of Vinalium**

## Example Data Problem

The What3Words app is a widely promoted system that is used for location sharing. It works by dividing the world up into three-by-three-metre squares. This app has been adopted by the emergency services as an easy way to pin-point locations, however, it has been found to use some homophones (similar sounding words), which, if incorrectly communicated or heard, will lead to the emergency response going to the wrong location, especially when the phone signal is poor, or someone is speaking in distress.




Poster sponsored by

**"There is a forest of data and we need to create a path through"**
Tom Adams

SCSC-127I

# Multi-Core Working Group

## Vision

To better understand the safety implications of multi- and manycore implementations, including characterisation and bounding of interference, impacts on timing and how to recover from errors.

To produce an SCSC publication that provides guidance on assuring the safety of multi- and manycore implementations.

## Input


EASA

AMC 20-193

AMC 20-193 Use of multi-core processors

1. Purpose

1.1 This AMC describes an acceptable means, but not the only means, for showing compliance with the applicable airworthiness specifications for aspects related to multi-core processors (MCPs) contained in airborne systems and equipment used in product certification or ETSO authorisations. Compliance with this AMC is not mandatory, and an applicant may elect to use an alternative means of compliance. However, this alternative means must fully address the airworthiness concern associated with unacceptably reduced level of safety, and the applicability the Agency's acceptance of MCP architectures.

1.2 This AMC provides objectives for the demonstration of compliance with the applicable airworthiness specifications for airborne systems and equipment that contain MCPs, according to the applicability in Section 2 of this AMC.

2. Applicability

2.1 This AMC may be used by applicants, design approval holders, and developers of airborne systems and equipment, which contain MCPs, to be installed on type-certified aircraft, engines, and propellers. This also includes developers of ETSO articles.

This AMC does not apply when the MCP is one of the following:

- a processor that is not intended to execute airborne software, but only airborne software loaded by the OEM or off the hardware item that contains the MCP (i.e. A, B, or C). The classification of concern is verified through the applicable airborne equipment Functional Hazard Assessment (FHA).
- This AMC does not apply when the MCP is one of Level D or E.

If an applicant modifies the core of the MCP (such as by activating one or more additional cores or adding software of DAL A, B, or C), then the applicant should reassess the applicability of this AMC.

Section 3.3 of this AMC describes the objectives that apply according to the assigned DAL (A, B, or C) of the installed software or of the hardware item that contains the MCP.

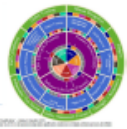
2.2 Aspects not addressed by this AMC

The following aspects are not covered by this AMC. This does not constitute an exemption, i.e. the objectives of this AMC are still applicable if an applicant has such a technique.

- any application-specific usage of these features should meet the intent that was used to show that the behaviour of the MCP is not affected, and determinism is still guaranteed.

In their planning activities, the applicant should present the methods employed to cover these aspects, and satisfy the objectives of this AMC, or show compliance with the applicable airworthiness specifications if they propose an alternative to this AMC or part of it.

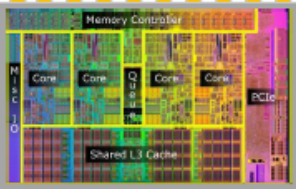
Input



## Activities




## Context

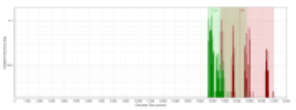



Challenges:

- Sharing of IP
- No usable case studies
- Significantly more complexity
- Solution dependent on specific hardware knowledge

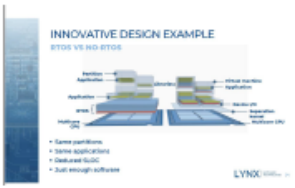



L2R Multicore Interference



Using Virtualization






Chairs: Louise Harney and Lee Jacques
35 Members
7 Sub Groups

MCWG-G1: Ontology
Michael Ward

MCWG-G2: Literature Review
Louise Harney

MCWG-G3: Review of Existing Standards and Guidance
Guillem Bernat

MCWG-G4: Safety Argument and Assessment
Iain Bate

MCWG-G5: Overarching Objectives
Lee Jacques

MCWG-G6: Processes for Meeting Standards and Guidance
Dave Speed

MCWG-G7: Case Studies
No Lead

THE SYSTEMS SAFETY COMMUNITY AND CLUB
SCSC
THE FUTURE OF SAFER SYSTEMS
SAFETY CRITICAL SYSTEMS CLUB

# Systems Approach to Safety of the Environment Working Group

https://scsc.uk/ge

## Mission Statement

***To produce clear guidance on how engineered systems should be developed and managed throughout their entire lifecycle so as to preserve, protect and enhance the environment***

### Why are we doing this?

- Systems safety and environmental safety have real and important differences between them in traditional safety-critical industry sectors (aviation, rail, etc.)
- Our idea is to apply a Systems Safety approach to systems which are embedded within the natural environment
- We are focussing on harm to the environment, which has a wider and more-encompassing basis than traditional safety engineering



### The Problem Space

- Systems effects have contributed to major accidents (Deep Water Horizon, Bhopal, etc).
- The impact of systems is not sufficiently addressed current environmental safety management practices and standards
  - These tend to be prescriptive and regulation-based
- Knock-on effects and emergent behaviours of systems and eco-systems are not currently assessed and managed from a multi-dimensional perspective
  - Introducing systems thinking, tools and techniques could help manage environmental risks due to complexity
- There are business, societal benefits and wider environmental benefits, in terms of reduced harm, reduced liabilities and improved business efficiency, if we improve management of systems risk related to safety of the environment
- Current commercial and legal penalties do not appear to mitigate environmental risk (e.g. dumping of sewage after recent heavy rains)
- Could an assurance case approach with claims and arguments be more appropriate than rules and regulations in isolation?

### Target Audience

- Regulators
- System developers
- Development project managers
- Operators / maintainers
- Procurement / commercial
- Government/Local Government
- Environmental Groups
- Those impacted by systems

### Applicability - Categories of Systems

- Fixed 'assets'
- Mobile systems
- Systems moved to new environments (not previously envisaged)

### National / Cultural Factors

- Local sensitivities
- Historical sites
- National / Cultural / Religious issues
- Cultural awareness across nationalities

### What can SCSC do better?

### Topic Speakers

### Grow the WG with appropriate expertise

### Develop Initial Issue of the Guidance

### Release Initial Issue of the Guidance




### Differences from Safety Engineering - 1

- **Harm not immediately obvious**: can be long-lasting, difficult to assess; and materialise over a longer time frame
- **Broader boundaries**: systems can have a wide scope and more affected parties (e.g. pollution moving globally in the air and ocean)
- **Multiple stakeholders**: competing stakeholders demands from commercial, governmental, national, local, campaign groups, NGO and individual interests
- **Increased complexity**: more interactions and relationships between elements (and stakeholders)
- **Ambiguity of scope**: systems boundaries may be 'leaky' or less well-defined

### Differences from Safety Engineering - 2

- **Trade-offs exist over multiple aspects**: how should different benefits be weighted? When should change be made to minimise environmental impact?
- **More unknowns and 'Dark Consequences'**: we don't know future effects will be in the future, particularly on human health
- **Long-term aggregation of harms**: cumulative effects and combinations can be significant and aren't often covered by normal risk assessment; 'Ecological Ceilings'
- **Wider societal aspects**: business, legal and ethical issues, e.g. shipping domestic waste to other countries

### SASEWG Approach

- Introduce how to apply the safety toolbox to environmental issues: safety analyses, formal assurance, safety case paradigm, etc.
- Explain how safety-critical systems thinking can support environmental and sustainability concepts, e.g. doughnut economics, circular economy
- Consider how different risk models may be appropriate, e.g. Black Swans, Grey Swans, Perfect Storms, Dragon King events, etc.
- Develop alternative decision-making heuristics beyond ALARP.



Produce new guidance to improve the situation
Feed into other SCSC groups to influence direction

### Differences from Safety Engineering - 3

- **Regulatory focus:** Existing environmental safety methods tend to be regulation and compliance-based
- **Need to be 'holistic':** a Systems-of-Systems approach and whole systems view is needed
- **Need for a long-term cumulative approach**: "My single-use plastic straw won't break the camel's back..."
- **Wider set of effects:** include emergent behaviours, chains of events and knock-on effects
- **Aim to improve**: include opportunity management to enhance or restore ecosystem, not just avoid accidents; include mental health, wellbeing, etc. as well as natural environment

### Differences from Safety Engineering - 4

- **"Tipping points' in complex systems** (Dragon King Events), e.g. population gets too small to be viable
- **Intentional impacts**: many environmental aspects do not require a failure to have an impact
- **Next-generation problem**: long-term issues and policies that impact beyond the current planning horizons, e.g. Climate change / Net Zero
- Role of insurance versus avoidance of impacts; fines, licenses, trading schemes
- Dominant culture of not being blamed / found out
- Need to decide when change to a better system is viable

### Guidance Document Planned Contents

1. **Introduction**
   a. Mission Statement
   b. Rationale
      i. Statement of Problem
      ii. Differences with traditional system safety
   c. Applicability
   d. Definitions
   e. Intended Audience
   f. Existing Work
      i. MOD, UK, EU, ISO, etc.
2. **Risk Assessment Approaches**
   a. Preliminary Environmental Hazard/Impact Identification
   b. Evaluation Schemes
   c. Boundaries / Screening & Scoping
   d. Coverage (Impacts, Impacted groups, etc)
   e. Accumulation / Aggregation of harms
   f. Contribution to wider harms
   g. Unknowns
      I. Categories of Events: Black Swan, Perfect Storm, Dragon Kings, etc
      II. Dark Consequences, etc
      III. Non-linear Effects, Feedback Loops & Emergent Behaviours
   h. Improvements & Enhancements
   i. Priorities
3. **Analyses**
   a. Models
   b. Adaptation of existing analyses (STPA, FRAM, etc)
   c. Causality – need domain knowledge
   d. New analyses
   e. Tools / Techniques
   f. Supporting Data
   g. Proportionality (have to know when to stop)
4. **Step by Step Methodologies**
   a. New systems
      I. Lifecycles
   b. Changes to systems
      I. Lifecycles
      II. Additions
      III. Removal and Disposal
   c. Changes to Context
5. **Trade-Off Approaches & Techniques**
   a. Principles and Weightings
   b. Included / Excluded
   c. Limits
   d. Business Case Inputs & Oppositions
   e. Obsolescence vs. efficiency
6. **Other Factors**
   a. Local and Cultural Sensitivities
   b. National/International Dimensions
   c. Legal & Compliance Aspects
   d. Moral and Ethical Considerations
   e. Personal Accountability & Justifications
7. **Making Justifications**
   a. Introduction
   b. Argument pattern templates
   c. Systems Approach Environmental Safety Case (SAESC)
   d. Systems Approach Environmental Safety Case Report (SAESCR)
   e. Feed through into SMS/EMS
8. **Reviewing & Auditing**
   a. Risk-based & Cover Systems Aspects
9. **Examples / Case Studies**
10. **Relationship to Other Work / Standards**
11. **Further Work**

**Are you passionate about the Environment?**
**Can you provide expertise in the planned content of the guidance?**
**Would you like to join the working group or attend as a guest speaker to the forum?**
**If so, get in touch: james.inge@scsc.uk**



# Safety of Autonomous Systems Working Group

SASWG

SCSC

## Vision

***To produce clear guidance on how autonomous systems and autonomy technologies should be managed in a safety related context, throughout the lifecycle, in a way that is tightly focused on challenges unique to autonomy.***

## Who we are

SASWG started in 2017 to tackle the safety issues associated with achieving the positive potential of autonomous systems (AS).

Members represent avionics, automotive, defense, maritime, medical, nuclear, regulators, practitioners, academics etc.

Meet every few months (online and hybrid) to discuss guidance and share experiences.

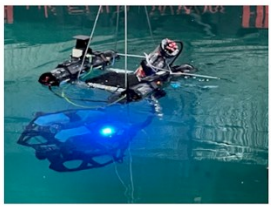
## Safety Assurance Objectives for Autonomous Systems




- SASWG guidance contains 45 objectives that should be addressed to achieve demonstrably safe autonomous systems
- Considers issues at *computation/algorithm, architectural* and *platform* level
- Focus is on Machine Learning (ML) elements of AS
- Guidance available via SCSC website and Amazon (TBC)

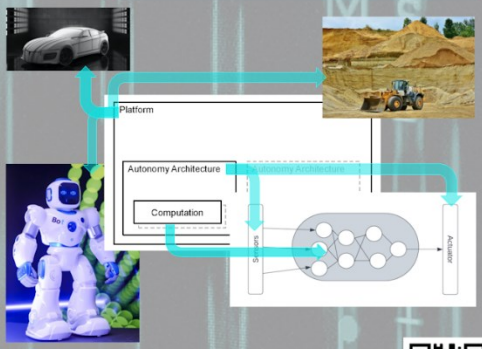
## Case study

- SASWG guidance applied to autonomous vehicle for nuclear decommissioning by D-RisQ
- Used to support safety summary to client and regulator



## How it works

- Example objectives:
  - COM1-1: Data is acquired and controlled appropriately
  - ARC1-2: Operational inputs inconsistent with training inputs are tolerated
  - PLT1-1: All aspects of platform behaviour that are achieved using autonomy-enabling techniques are justified
- Evidence should be provided to support each objective, or objective justified as not relevant
- Creates context specific safety justification for ML elements




*This poster is kindly sponsored by*

ATKINS
Member of the SNC-Lavalin Group

HEMRAJ
GLOBAL EXPERTS IN FUNCTIONAL SAFETY DESIGN FOR THE AUTOMOTIVE, AEROSPACE, & DEFENCE INDUSTRIES

D-RisQ
SOFTWARE SYSTEMS

**Contact us:**

Webpage https://scsc.uk/ga

Email: philippa.ryan@york.ac.uk (Chair)

Background photo by Markus Spiske freegreatpicture.com from Pexels https://www.pexels.com/



Assurance Wrappers 'O'




Utility Services



Plant Services



Location Services



Health Services

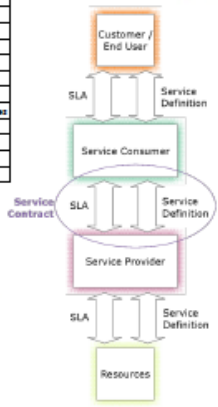


Service Contract

Off-Shore Services



| Service Characteristics | |
|---|---|
| C1 | Services are provided for the duration of a service contract |
| C2 | Services are often designed to meet the needs of a broad range of consumer needs |
| C3 | Services are likely to be used by more than one consumer |
| C4 | Services are implemented through a combination of elements |
| C5 | Services may be designed without recognition of the full context of use |
| C6 | Service implementation details may not be visible to the consumer |
| C7 | Service implementation may change frequently and significantly |
| C8 | Services may include or provide features for the maintenance or monitoring of the service or other services |
| C9 | Services are often multi-layered |
| C10 | Services may be created rapidly |
| C11 | Services may change size and scope quickly |
| C12 | Services may be provided from frequently changing multiple locations |




Aviation Services



Highway Services



Railway Services



Marine Services



Defence Services





## Service Assurance Principles

The intent of the service assurance requirements shall be maintained through the service definitions, service levels, the service architecture and the agreements made at service interfaces

Unintended behaviours of the service-based solution shall be identified, assessed and managed

These principles shall be established and maintained throughout the lifetime of the SBS, resilient to all changes and re-purposing

1 Requirements shall be defined to address the service-based solution's (SBS) contribution to both desirable and undesirable behaviours

2

3 Service assurance requirements shall be satisfied

4

5 The confidence established in addressing these principles shall be commensurate with the level of risk posed by the service-based solution

6

# Service Assurance Working Group

*"To produce clear and practical guidance on how services should be managed in a safety related context, to reflect emerging best practice"*

Members: 37
Meetings: 53

https://scsc.uk/gs
contact: kevin.king@scsc.uk




The Systems Safety Community and Club

| Level of Service Assurance | Definition (Service Consumer View) |
|---|---|
| LSA 0 | No safety aspects at present in service so no objectives assigned to services assessed to be at LSA0 |
| LSA 1 | Minor safety aspects with little impact of failures (minor injury possible but unlikely) |
| LSA 2 | Safety aspects with some impact of failures (several injuries possible) |
| LSA 3 | Significant safety aspects with service with major impact (could indirectly lead to multiple injuries or a single death) |
| LSA 4 | Service is safety-critical: service failures could have catastrophic impact (could directly lead to multiple deaths) |

SCSC

# SCSWG

## Safer Complex Systems Working Group

Meetings 7 - Members 34

*To produce practical guidance on developing, managing and assuring safety of complex systems throughout their lifecycle*

Systems are now too complex to sensibly reason about their safety…

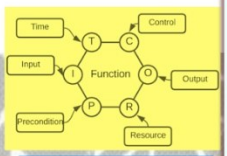
…we don't know what will happen when in operation with potential sophisticated and emergent behaviours…

…we can't produce compelling assurance cases…

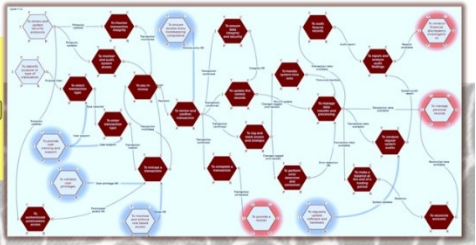
…we need better approaches and some guidance to help us build and operate such systems




- ✓ Build on IET and RAE work
- ✓ Produce generic Guidance
- ✓ Toolbox: analyses, safety case
- ✓ Systems of systems approach
- ✓ Include emergent properties




Evaluate analysis methods such as FRAM



Nats air traffic control fault: Experts reflect on three days of chaos



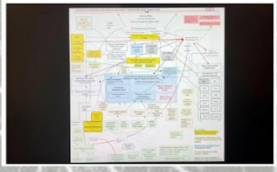
Learn from incidents with complexity issues

A system is never the sum of its parts; it's the product of their interaction - Ackoff

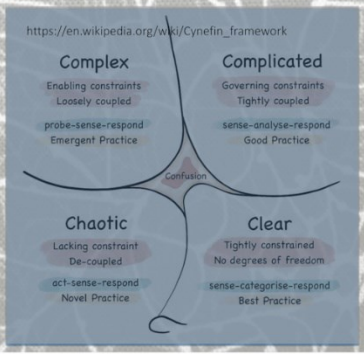
We need to assess risk, but this is a unsolved problem for complex critical systems; we need efficient and practical risk assessment

'A bowl of spaghetti' : Covid inquiry opens with flowchart on UK's pandemic planning

Over 100 bodies appear to have had a role in protecting Britain from deadly viruses, and Lady Hallett must untangle the mess






Snowden's CYNEFIN: we need more than system thinking… there are very different behaviours in predictable, but complicated systems, and truly complex systems that can behave unpredictably


Contact:
mike.parsons@scsc.uk
scsc.uk/gp

Security Informed Safety Working Group (SISWG)

# How do we manage the impact of (cyber)security on functional safety?

Stephen Bull, Nikita Johnson, Wendy Owen, Paul Hampton

contact: stephen.bull@scsc.uk

## The problem ...

- **(Cyber)security vulnerabilities can be exploited to compromise the safety of systems, causing catastrophic consequences!**
- **This is a complicated and growing problem as systems become more complex, interconnected and data-driven.**

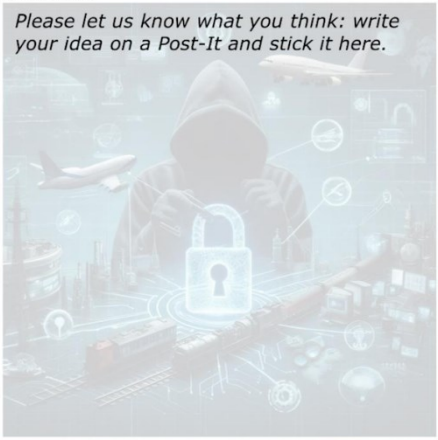
## The effects ...

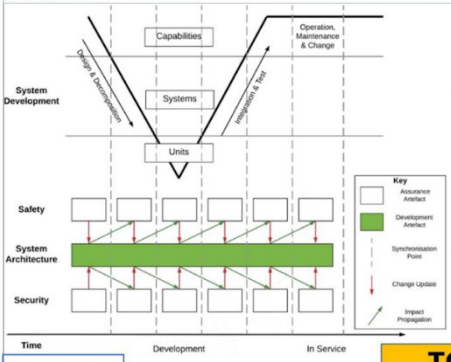






*Please let us know what you think: write your idea on a Post-It and stick it here.*






## Safety-Security assurance framework

- Separate assurance processes, sharing the right information with the right people at the right time.
- Loose coupling across disciplines
- Mutual understanding to facilitate synchronised solution evolution through the whole lifecycle.

## Framework Principles:

*Which do you find most challenging?*

Essential principles:
- Independent Co-assurance
- Synchronisation Points
- Defined Lifecycle Phases
- Risk Impact Propagation

Supporting principles:
- Model-based representation
- Communication by Artefacts
- Proportionality

For future consideration
- Assurance Surface
- Separation of Concerns

scsc.uk/gi

**TG1: Co-assurance lifecycle**

Differences in goals and approach make unified safety and security analysis difficult

**TG2: Risk assessment process**

**TG3: Common language**

TG3 material bridges the language gap between safety and security to aid understanding.

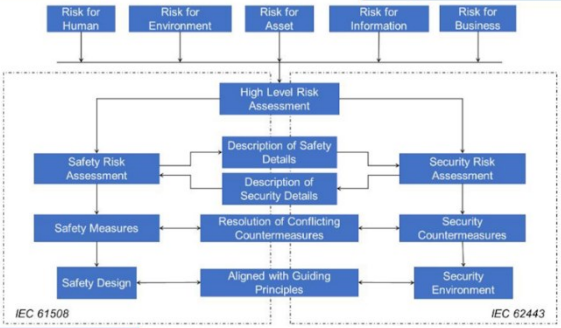



Illustration of a possible risk assessment process: TG2 will provide guidance on addressing security within safety processes.

**TG2: Key topics**

*Please add a sticky dot to show which is most important to you.*

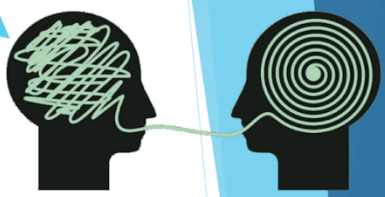
- Comparison of risk of harm from accidents with the risk from cyber attacks
- Inclusion of cyber risks in established safety techniques
- Assessing frequency of cyber risks
- Analysis of risks emerging during operation
- Analysis of cyber risk for legacy systems
- Balancing the needs of safety and security

Poster sponsored by Ebeni

# An Ontology of Risk Terminology for Safety and Security Engineering

SCSC Ontology Working Group - Website: https://scsc.uk/go - Email: ontology@scsc.uk

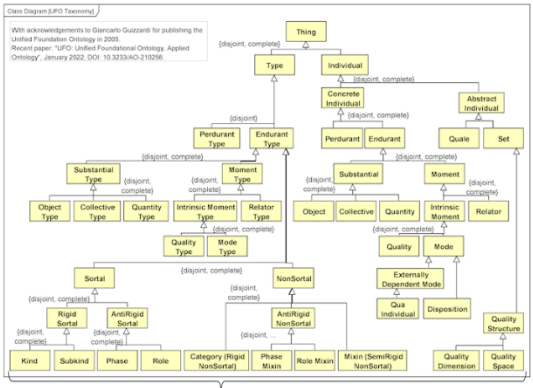


Variation in use and interpretation of language leads to misunderstanding and a lack of fully understood shared objectives; risk terminology originates from many sources including (worst of all) natural language.

**Ontologies** *conceptualise and formalise* the **World** by defining *concepts*, *terms*, and their *relationships*.

Examples of visual ontologies denoted as UML 2.5 class diagrams.

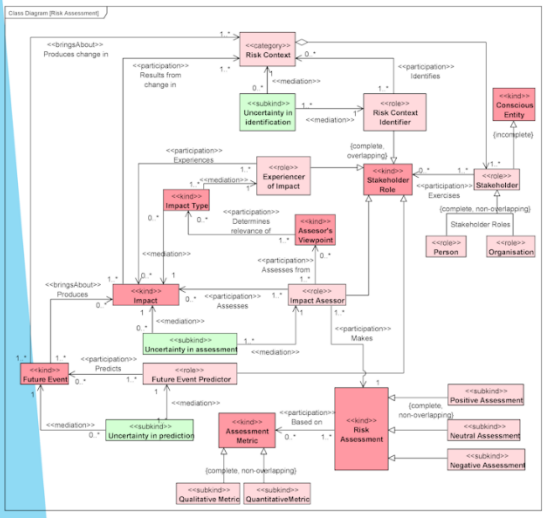


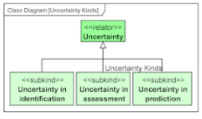

## The Ontology Working Group

OWG 2023 Objectives:

- Develop a clear, systematic, and unambiguous model of risk concepts and terminology and their relationships.
- Adoption of the Unified Foundation Ontology (UFO).
- Use of visual ontology modelling methods; UML and OntoUML profile.
- Liaison with ISO & IEC joint working group on Risk.

UFO Taxonomy derivative terms define the OntoUML stereotypes. Relationships infer modelling rules for the use of the stereotypes.

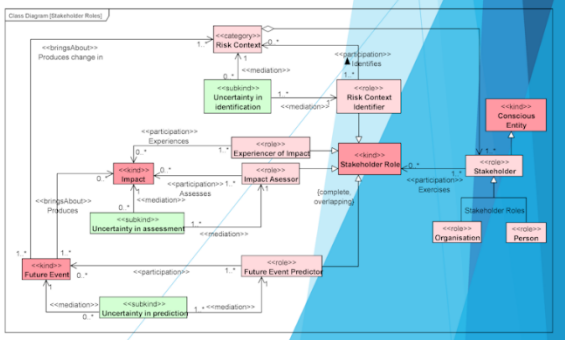




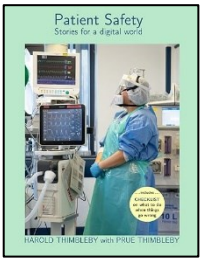


Example application of OntoUML for risk concepts




BlackBerry. QNX.
sponsored this poster

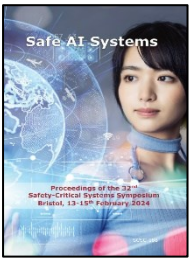
# Recent Safety Publications



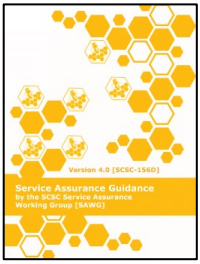

Patient Safety – Stories for a digital world
www.amazon.co.uk/dp/1399975420



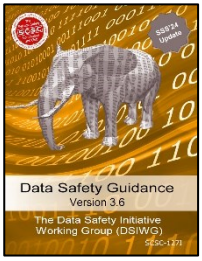

Safety Assurance Objectives for Autonomous Systems.
scsc.uk/scsc-153C



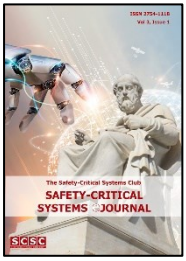

Proceedings of the 32nd Safety-Critical Systems Symposium.
scsc.uk/scsc-188




Service Assurance Guidance Version 4.0. Jan 2024
scsc.uk/scsc-156D



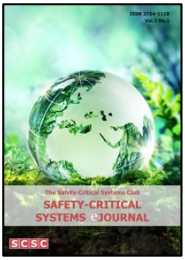

Guidance on the management of data safety risks (V3.6)
scsc.uk/scsc-127I



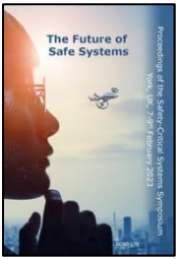

Safety-Critical Systems eJournal vol.3 no.1
Winter Issue 2023
scsc.uk/scsc-191




Safety Culture Working Group Position Paper
Edition 1. Dec 2023
scsc.uk/scsc-189




Safety-Critical Systems eJournal vol.2 no.2
Summer Issue 2023
scsc.uk/scsc-186




Proceedings of the 31st Safety-Critical Systems Symposium.
scsc.uk/scsc-179

# 60 Seconds with … Carmen Carlan



**Carmen Carlan has been working on topics related to the safe development of complex systems for nine years. Currently, she works as a system safety engineer at Edge Case Research (ECR), where she helps commercial clients solve the challenges they face related to the safety assurance of AVs in the automotive domain.**

**She also works on publicly funded research projects, keeping in touch with the state of the art in the domain of trustworthy artificial intelligence.**

**Further, Carmen occupies the position of Director of Safety Products, overviewing the development and management of ECR's safety products, such as the Open Autonomy Safety Case Framework, the nLoop platform, and templates for various safety work products in compliance with standards such as ISO 26262, ISO 21448, or MIL-STD 882.**

**Previously, she worked on her dissertation on developing and maintaining model-based safety cases. Carmen has also recently joined the SCSC Steering Group.**

## What first attracted you to working in the field of System Safety?

The idea that working in this field allows me to implicitly help people, by protecting them from accidents.

## What aspect of your career are you most proud of?

The fact that I managed to transfer part of my research to industry projects.

## Tell us one interesting unusual fact about yourself!

I am a Christian Orthodox and, personally, I think that being a Christian also helps with being a good system safety engineer, as Christianity is all about no discrimination, which is especially now relevant, given the usage of Artificial Intelligence (AI) in safety-critical applications, and trying to not harm anyone.

## What advice would you give to yourself age 12?

Enjoy the carefree life and use the time to read as much as you can. I am an avid reader, but these days I only get to read books/articles for my job, and I rarely find the time for reading a novel, for example.

## What future changes would you like to see in the field of System Safety?

I would like to see more courses on system safety engineering in German universities. Also, I would like to see more interaction between AI experts, safety engineers, security engineers, and people working on ethics. Further, I think there is a need to soon have experts on both AI and safety engineering. Last but not least, I want to see that companies stop treating safety as an afterthought.

## What's your most favourite quote or motto?

"Both religion and science need for their activities the belief in God, and moreover God stands for the former in the beginning, and for the latter at the end of the whole thinking. For the former, God represents the basis, for the latter – the crown of any reasoning concerning the world-view." – *Max Planck, Religion und Naturwissenschaft, Leipzig: Johann Ambrosius Barth Verlag, 1958, 27*

## If you could learn to do anything, what would it be?

Currently, I think I have to learn more about AI.

## If you could be any fictional character, who would you choose?

Sherlock Holmes. I love his perseverance and determination in solving complicated problems. However, for now, I'll stick to solving problems in system safety engineering.

## Artificial Intelligence - a positive revolution or are we all doomed!?

I think we just need to wait and see (:

# Connect

## The Newsletter and eJournal

Do you have a topic you'd like to share with the systems safety community? Perhaps an interesting area of research or project work you've been involved in, some new developments you'd like to share, or perhaps you would simply like to express your views and opinions of current issues and events. There are now two publishing vehicles for content – shorter, more informal content, can be published in the Newsletter with longer, more technical peer-reviewed material more suitable for the eJournal. If you are interested in submitting content, then get in touch with Paul Hampton for Newsletter articles: paul.hampton@scsc.uk or John Spriggs for eJournal papers: john.spriggs@scsc.uk

## The SCSC Website

Visit the Club's website thescsc.org for more details of the Safety-Critical Systems Club including past newsletters, details of how to get involved in working groups and joining information for the various forthcoming events.

## Facebook

Follow the Safety-Critical Systems Club on its very own Facebook page.

www.facebook.com/SafetyClubUK

## X/Twitter

Follow the Safety-Critical Systems Club's X/Twitter feed for brief updates on the club and events: @SafetyClubUK

## LinkedIn

You can find the club on LinkedIn. Search for the Safety-Critical Systems Club or use the following link:

www.linkedin.com/groups/3752227

## Advertising

Do you have a product, service or event you would like to advertise in the Newsletter? The SCSC Newsletter can reach out to 1,000's of individuals involved in Systems Safety and so is the perfect medium for engaging with the community. For prices and further details, please get in touch with the Newsletter Editor.

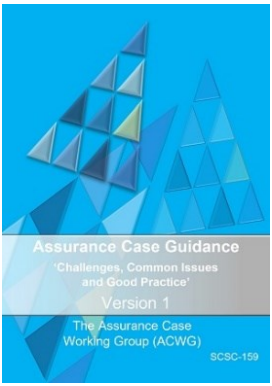
# SCSC Working Groups

The Safety-Critical Systems Club is committed to supporting the activities of working groups for areas of special interest to club members. The purpose of these groups is to share industry best practice, establish suitable work and research programmes, develop industry guidance documents and influence the development of standards.

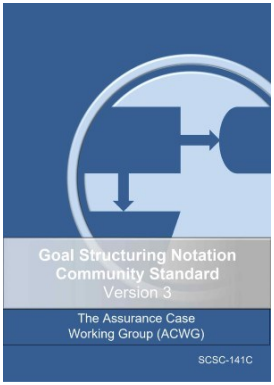
## Assurance Cases

The Assurance Cases Working Group (ACWG) has been established to provide guidance on all aspects of assurance cases including construction, review and maintenance. The ACWG will:

- Be broader than safety, and will address interaction and conflict between related topics
- Address aspects such as proportionality, rationale behind the guidance, focus on risk, confidence and conformance
- Consider the role of the counter-argument and evidence and the treatment of potential bias in arguments




In Aug 2021, the group published v1.0 of the Assurance Case Guidance: scsc.uk/scsc-159




One of the working group's activities is the maintenance of the Goal Structuring Notation (GSN) Community standard.

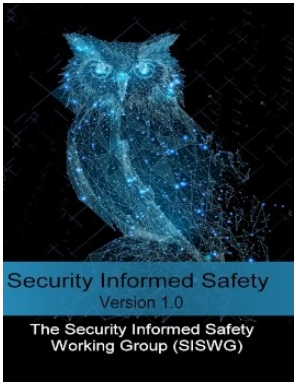
See scsc.uk/gsn for further details.

In May 2021, the group published v3.0 of the standard: scsc.uk/scsc-141C

**Lead Phil Williams** phil.williams@scsc.uk

# SCSC Working Groups

## Security Informed Safety




The Security Informed Safety Working Group (SISWG) aims to capture cross-domain best practice to help engineers find the 'wood through the trees' with all the different security standards, their implication and integration with safety design principles to aid the design and protection of secure safety-critical systems and systems with a safety implication.

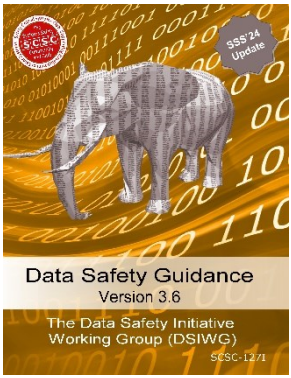
The working group aims to produce clear and current guidance on methods to design and protect safety-related and safety-critical systems in a way that reflects prevailing and emerging best practice.

The guidance will allow safety, security and other stakeholders to navigate the different security standards, understand their applicability and their integration with safety principles, and ultimately aid the design and protection of secure safety-related and safety-critical systems.

**Lead** **Stephen Bull** stephen.bull@scsc.uk

## Data Safety Initiative

Data in safety-related systems is not sufficiently addressed in current safety management practices and standards.




It is acknowledged that data has been a contributing factor in several incidents and accidents to date and there is foreseeable harm that can arise from Machine Learning and Large Language Models' use of data by Artificial Intelligence (AI) systems that are subject to issues of biasing, interpretation and, arguably, falsification. There are clear business and societal benefits, in terms of reduced harm, reduced commercial liabilities and improved business efficiencies, in investigating and addressing outstanding challenges related to safety of data.

The Data Safety Initiative Working Group (DSIWG) aims to have clear guidance on how data (as distinct from the software and hardware) should be managed in a safety-related context, which will reflect emerging best practice.

An update to the guidance (v3.6) was published in Jan 2024: scsc.uk/scsc-127I

**Lead** **Mike Parsons** mike.parsons@scsc.uk

# SCSC Working Groups

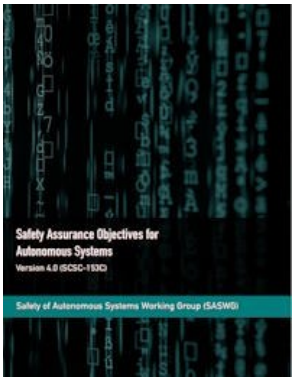
## Safe AI

The SCSC Safe AI Working Group (SAIWG) aims to capture cross-domain best practice and guidance on key topics within the design, evaluation, assurance, and approval of safety systems that use or are developed using AI, bringing together emerging standards and key results from the incredible amount of research being conducted into AI safety.

The working group was kicked-off at SSS'24 and is led by Alan Simpson. The SAIWG will conduct regular meetings, workshops, and publications to share knowledge and experience on various topics related to AI and safety systems, such as coordination of safety with other disciplines, evaluation of risk, and mapping of terminology and language.

**Lead** **Alan Simpson** alan.simpson@ebeni.com

## Safety of Autonomous Systems




The specific safety challenges of autonomous systems and the technologies that enable autonomy are not adequately addressed by current safety management practices and standards.

It is clear that autonomous systems can introduce many new paths to accidents, and that autonomous system technologies may not be practical to analyse adequately using accepted current practice. Whilst there are differences in detail, and standards, between domains many of the underlying challenges appear similar and it is likely that common approaches to core problems will prove possible.

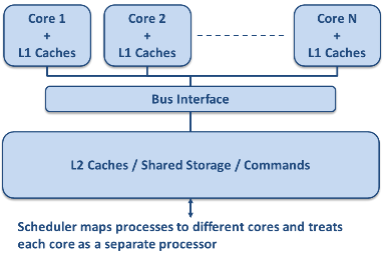
The Safety of Autonomous Systems Working Group (SASWG) aims to produce clear guidance on how autonomous systems and autonomy technologies should be managed in a safety-related context, in a way that reflects emerging best practice.

The group will work closely with the new Safe AI Working Group (SAIWG) as AI is used extensively in Autonomous Systems.

The group published v4 of its guidance Safety Assurance Objectives for Autonomous Systems, in Feb 2024 scsc.uk/scsc-153C

**Lead** **Philippa Ryan** pmrc@adelard.com

## Multi- and Manycore Safety




It is becoming harder and harder to source single-core devices and there is a growing need for increased processing capability with a smaller physical footprint in all applications. Devices with multiple cores can perform many processes at once, meaning it is difficult to establish (with sufficient evidence) whether or not these processes can be relied upon for safety-related purposes.

Parallel processes need to access the same shared resources, including memory, cache and external interfaces, so they may contend for the same resources. Resource contention is a source of interference which can prevent or disrupt completion of the processes, meaning it is difficult to know with a defined uncertainty the maximum time each process will take to complete (Worst Case Execution Time, WCET) or whether the data stored in shared memory has been altered by other processes.

The Multi- and Manycore Safety Working Group (MCWG) has been established to explore the future ways of assuring the safety of multi- and manycore implementations.

**Lead** **Lee Jacques** Lee.Jacques@leonardocompany.com

## Safer Complex Systems



The Safer Complex Systems Working Group (SCSWG) builds on the IET/RAE work already done in this area. It is recognised that the RAE work is ongoing and collaboration is encouraged. The group's mission is to produce practical guidance on developing, managing and assuring complex systems throughout their lifecycle (so as to achieve and justify their safety).

The following provides a statement of the problem:

- It is acknowledged that complex systems are becoming more prevalent with more opportunities to cause harm
- There are also new complex systems arising from using combinations of existing systems and services which are then used for safety purposes
- Complex systems are not sufficiently addressed in current safety management practices and standards
- In particular, complex interactions and emergent behaviours are not currently assessed and managed sufficiently
- There could be benefit in developing new analysis, tools and techniques to manage complex system risks
- There are clear business and societal benefits, in terms of reduced harm, reduced liabilities and improved business efficiencies, in improved management of complex systems risk

Contact the group lead if you would like to attend future meetings or find out more about the working group.

**Acting Lead** **Mike Parsons** mike.parsons@scsc.uk

# SCSC Working Groups

## Service Assurance

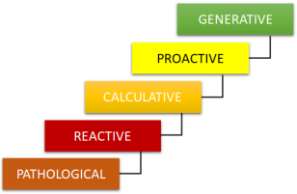



Risks presented by safety-related services are rarely explicitly recognised or addressed in current safety management practices, guidelines and standards. It is likely that service (as distinct from system) failures have led to safety incidents and accidents, but this has not always been recognised. The Service Assurance Working Group (SAWG) has been set up to produce clear and practical guidance on how services should be managed in a safety-related context, to reflect emerging best practice.

The group published guidance v4.0 in Jan 2024: scsc.uk/scsc-156D

**Lead** **Kevin King** kevin.king@baesystems.com

## SCSC Safety Culture




The Safety Culture Working Group (SCWG) has been established to provide guidance on creating and maintaining an effective safety culture. The group seeks to improve safety culture in safety-critical organisations focussed on product and functional safety, by sharing examples and latest approaches collated from real-life case studies.

Meetings provide an opportunity to discuss any particular aspects attendees are interested in taking forward, and to help set future directions for the group.

In Dec 2023, the group published a position paper for assessing and managing safety culture. https://scsc.uk/r189:1

**Lead** **Michael Wright** michael.wright@greenstreet.co.uk

## Systems Approach to Safety of the Environment



The Systems Approach to Safety of the Environment Working Group (SASEWG) is a new group intending to apply Systems Safety practices to systems that are embedded within the natural environment, while focussing on that environment.

The group aims to produce clear guidance on how engineered systems should be developed and managed throughout their entire lifecycle so as to preserve, protect and enhance the environment.

Please get in touch with the working group lead if you would like to join or find out more about this group.

**Lead** **James Inge** james.inge@scsc.uk

# SCSC Membership

The SCSC provides a range of services to the System Safety community including seminars, tutorials, leadership events, specialist topic working groups, the annual symposium and a comprehensive body of publications. Membership brings many valuable benefits such as free access to online events, the SCSC Newsletter and access to presentations and other resources from events.

## Individual Membership

To become an individual member of the SCSC please register on the SCSC website using the icon at the top right of any page and select "Register". Complete and save your account registration and then verify your email address. Once registered and logged in click the link "why not join the SCSC..." inviting you to become a member at the top right of the page or select "Pay membership" from the icon.

Individual membership can be paid online using a credit/debit card through our secure payment partner Realex Global Payments or contact Alex King for other payment methods. For student or retired member rates please contact Alex King to get your account status changed.

## Corporate Membership

Your company contact with the SCSC should arrange the membership and any renewals for your organisation. To join as a member covered by a corporate membership, register as per the instructions for an individual member and then contact Alex King to confirm your affiliation.

## Renewing Membership

You should be notified by email when your membership is almost expired or shortly after it has expired. These notifications will contain a link to the online renewal page or you will be able to renew when logging onto the website through the 'click to renew' link.

## Membership Fees

The following fees are applicable for new and renewing members:

- 1 year Individual Membership: £145
- 2 year Membership: 13% discount: £250
- 3 year Membership: 19% discount: £350
- 1 year SFI Membership: FREE for first year, £35 for years 2 & 3
- 1 year Membership, retired member rate: £35
- For Corporate Membership discounts contact Alex King

A one-month Publication Pass is also available for £15. This allows access to all SCSC website publications in a particular calendar month.

Contact Alex King using office@scsc.uk

# The SCSC Steering Group


Tom Anderson
*Honorary member*


Stephen Bull
stephen.bull@scsc.uk


Dewi Daniels
dewi.daniels@scsc.uk


Jane Fenn
jane.fenn@scsc.uk


Louise Harney
louise.harney@scsc.uk


Brian Jepson
brian.jepson@scsc.uk


Tim Kelly
*Honorary member*


Mark Nicholson
mark.nicholson@scsc.uk


Wendy Owen
wendy.owen@scsc.uk


Davy Pissoort
davy.pissoort@scsc.uk


Roger Rivett
roger.rivett@scsc.uk


Sean White
sean.white@scsc.uk


Robin Bloomfield
*Honorary member*


Carmen Carlan
carmen.carlan@scsc.uk


Dai Davis
*Honorary member*


Paul Hampton
paul.hampton@scsc.uk


James Inge
james.inge@scsc.uk


Graham Jolliffe
*Honorary member*


Alex King
alex.king@scsc.uk


Yvonne Oakshott
yvonne.oakshott@scsc.uk


Mike Parsons
mike.parsons@scsc.uk


Felix Redmill
*Honorary member*


John Spriggs
john.spriggs@scsc.uk


Phil Williams
*Honorary member*

# Club Positions

The current and previous (marked in italics) holders of club positions are as follows:

## Managing Director

**Mike Parsons 2019-**

*Tim Kelly 2016-2019*
*Tom Anderson 1991-2016*

## Programme & Events Coordinator

**Mike Parsons 2014-**

*Chris Dale 2008-2014*
*Felix Redmill 1991-2008*

## Newsletter Editor

**Paul Hampton 2019-**

*Katrina Attwood 2016-2019*
*Felix Redmill 1991-2016*

## eJournal Editor

**John Spriggs 2021-**

## Diversity, Equity and Inclusion (DE&I) Lead

**Wendy Owen 2023-**

## Steering Group Chair

**Roger Rivett 2019-**

*Graham Jolliffe 2014-2019*
*Brian Jepson 2007-2014*
*Bob Malcolm 1991-2007*

## Manager

**Alex King 2019-**

## Honorary Solicitor

**Dai Davis 2022-**

## University of York Coordinator

**Mark Nicholson 2019-**

## Website Editor

**Brian Jepson 2004-**

## Administrator

**Alex King 2016-**

*Joan Atkinson 1991-2016*

## Safety Futures Initiative Leads

**Khadijah Khatun 2024-**
*Zoe Garstang 2019-2024*
*Nikita Johnson 2019-2021, 2023-2024*

# Calendar

**May '24**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| 27 | 28 | 29 | 30 | 31 | | |

**June '24**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| | | | | | 1 | 2 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| 24 | 25 | 26 | 27 | 28 | 29 | 30 |

**July '24**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| 29 | 30 | 31 | | | | |

**August '24**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| 26 | 27 | 28 | 29 | 30 | 31 | |

**September '24**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| | | | | | | 1 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 |

**October '24**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 21 | 22 | 23 | 24 | 25 | 26 | 27 |
| 28 | 29 | 30 | 31 | | | |

**November '24**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| | | | | 1 | 2 | 3 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 25 | 26 | 27 | 28 | 29 | 30 | |

**December '24**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| | | | | | | 1 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 |

**January '25**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| 27 | 28 | 29 | 30 | 31 | | |

**February '25**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| | | | | | 1 | 2 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| 24 | 25 | 26 | 27 | 28 | | |

**March '25**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| | | | | | 1 | 2 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| 24 | 25 | 26 | 27 | 28 | 29 | 30 |

**April '25**

| M | T | W | T | F | S | S |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 21 | 22 | 23 | 24 | 25 | 26 | 27 |
| 28 | 29 | 30 | | | | |

# Events Diary



**21 May 2024**
SCSC Working Group
**Zoom 4pm-5pm**

**SCSC 84th Data Safety Initiative Working Group Meeting**
scsc.uk/e1127

**4-7 June 2024**
**Conference**
**Ferrara, Italy**

**Safe Process 2024: 12th IFAC.**
safeprocess2024.eu

**11-14 June 2024**
**Conference**
**Barcelona, Spain**

**Ada-Europe: 28th Int.**
ada-europe.org/conference2024

**12-13 June 2024**
**Conference**
**Tampere, Finland**

**Safety of Industrial Automated Systems**
automaatioseura.fi/sias2024

**23-27 June 2024**
**Conference**
**Cracow, Poland**

**34th European Safety and Reliability Conference**
esrel2024.com

**24-27 June 2024**
**Conference**
**Brisbane, Australia**

**The 54th Annual IEEE/IFIP International Conference**
dsn2024uq.github.io

**27 June 2024**
SCSC Seminar
**London, UK**

**Developing Safe AI Systems**
scsc.uk/e1100

**8 July 2024**
**Workshop**
**Enschede, Netherlands**

**Sensei 2024: Safety-Security Interaction**
sites.google.com/view/sensei-2024

**17 Sept 2024**
**Workshop**
**Florence, Italy**

**WAISE 2024: 7th Int. Workshop on Artificial Intelligence Safety Engineering**
www.waise.org

**17 Sept 2024**
**Workshop**
**Florence, Italy**

**SASSUR 2024: 11th Int. Workshop on Next Generation of System Assurance Approaches for Critical Systems**
sites.google.com/view/sassur2024

**17-20 Sept 2024**
**Conference**
**Florence, Italy.**

**43rd International Conference on Computer Safety, Reliability and Security**
safecomp2024.unifi.it

**26 Sept 2024**
SCSC Seminar
**London, UK**

**Deployment, Operations and Maintenance of Safe AI Systems**
scsc.uk/e1133

**22 Oct 2024**
**Conference**
**Newport, UK**

**HISC 2024: High Integrity Software Conference**
www.his-conference.co.uk

**28-31 Oct 2024**
**Symposium**
**Tsukuba, Japan**

**ISSRE 2024: 35th IEEE Int. Symp. on Software Reliability Engineering**
issre.github.io/2024

**13-15 Nov 2024**
**Symposium**
**Osaka, Japan**

**PRDC 2024: 29th IEEE Pacific Rim Int. Symp. on Dependable Computing**
prdc.dependability.org/PRDC2024/

**20-22 Nov 2024**
**Conference**
**Catania, Sicily, Italy**

**ICSRS 2024: System Reliability and Safety**
www.icsrs.org

**28 November 2024**
SCSC Seminar
**Munich, Germany**

**Safe Autonomous Vehicles - the Good, the Bad and the Ugly**
scsc.uk/e1134

**26-29 Nov 2024**
**Symposium**
**Recife, Brazil**

**LADC 2024: Dependable and Secure Computing**
ladc.sbc.org.br/2024

**5-6 Dec 2024**
**Conference**
**Pisa, Italy**

**EAI SCS 2024: -Critical Software and Systems**
scs-conference.eai-conferences.org/2024

**4-6 Feb 2025**
SCSC Symposium
**York, UK**

**Safety Critical Systems Symposium SSS'25**
scsc.uk/e1099


thescsc.org/membership
Safety Critical Systems Club
Supporting Development of Safer Systems
The
Systems Safety
SCSC
Community
and Club