SCSC Data Safety Initiative – WG Meeting 81

6th December 2023, Zoom

Minutes

Attendees

Mike Parsons (MP) – Ebeni, Roland Rosier (RR) – TomTom, Martin Atkins (MA) – MCA, Dave Banham (DB) – Blackberry, Paul Hampton (PH) – CGI, Paolo Giuliani (PG) – EDF, Divya Atkins (DA) – MCA, Arch McKinley (AM) – NGA, Nick Hales (NH) – Consultant, Jennifer Ebbinge (JE) – TomTom, Oscar Slotosch (OS) – Validas, Brent Kimberley (BK) – Durham, Mark Templeton (MT) – QinetiQ, Paul Leftwich (PL) – Dstl.

Apologies

Mark Nicholson (MN) – University of York, Dave Murray (DM) - BAE Systems, Tim Rowe (TR) – Consultant.

Agenda

1. **Welcome**
2. **AI & Data**
3. **Poster for SSS'24**
4. **New Guidance Version for SSS'24**
5. **SITN (Data Safety in the News)**
6. **Update on Tooling**
7. **Actions**
8. **Next meeting**
9. **AOB**

NOTE: All comments or opinions in these notes are attributed only to individual attendees of the meeting, not to their respective organisations.

*[Note that actions are presented in the form **N.Mx** where **N** is the meeting number, **M** a reference number for the action raised in that meeting and **x** is an optional letter that differentiates related actions arising from the same discussion point].*

The meeting slides are available at: https://scsc.uk/file/gd/81st_DSIWG_Slides_v1-1643.pptx

1. Welcome

MP opened the meeting and welcomed those attending.

2. AI & Data

MP mentioned several news articles suggesting harmful recipes produced by a supermarket AI meal planner app: *“Pak ‘n’ Save’s Savey Meal-bot cheerfully created unappealing recipes when customers experimented with non-grocery household items”*:
https://www.theguardian.com/world/2023/aug/10/pak-n-save-savey-meal-bot-ai-app-malfunction-recipes
This was thought to be a good example of an AI-powered app that demonstrated ‘garbage-in, garbage-out’ as it was never intended that people would enter non-food items into the app. However, people will always do amusing (or dangerous) things with AI and this needs to be recognized. It seems that the app has now been improved:
https://www.trtworld.com/australia/new-zealand-supermarket-ai-meal-planner-app-suggests-poisonous-recipes-14417222

This led MP to experiment with ChatGPT 4:

User:

> *What are the risks of data in safety critical systems?*

ChatGPT4:

> *Data in safety-critical systems, such as those used in healthcare, aviation, nuclear power plants, and automotive safety systems, face several risks:*
>
> *Data Corruption: Corruption can occur due to software bugs, hardware failures, or external factors like electromagnetic interference. In safety-critical systems, corrupted data can lead to incorrect decisions or actions, with potentially catastrophic consequences.*
>
> *Data Loss: Loss of crucial data, whether through hardware failure, accidental deletion, or a catastrophic event, can render a safety-critical system unable to perform its essential functions.*
>
> *Cybersecurity Threats: Safety-critical systems are increasingly connected to other systems and the internet, making them vulnerable to cyber-attacks. This can include hacking, malware, and ransomware attacks, which could compromise the integrity or availability of critical data.*
>
> *Inaccurate Data: Inaccurate data, whether due to sensor errors, human input errors, or flawed data processing algorithms, can lead to incorrect outputs or decisions by the safety-critical system.*
>
> *Data Overload: Systems overwhelmed with data may fail to process it efficiently, leading to delays or incorrect prioritization of tasks. In a safety-critical context, this could result in a failure to respond appropriately to an emergency situation.*
>
> *Outdated or Obsolete Data: If a system relies on outdated information for decision-making, it may not account for recent changes or developments, leading to inappropriate or dangerous actions.*

> *Interoperability Issues: Safety-critical systems often need to communicate with other systems. Incompatibility in data formats or protocols can lead to misinterpretation of data or failure to communicate critical information.*
>
> *Privacy and Confidentiality Breaches: Particularly relevant in healthcare and personal data-driven systems, unauthorized access to sensitive data can lead to privacy violations and potential misuse of personal information.*
>
> *To mitigate these risks, rigorous testing, robust cybersecurity measures, data integrity checks, and regular system updates are essential in the design and maintenance of safety-critical systems.*

Which MP thought was a very introduction to the topic. It was noted that it didn't specifically make reference to the data safety guidance document or use its terminology.

There was some discussion on this.

DB noted that ChatGPT produced some entirely reasonable looking requirements for the C Standard Library function malloc(), although they were somewhat incomplete in their coverage of this function's required behaviour[1].

AM said that AI data needed metadata associated with it. He noted that trust in AI systems is a big issue.

RR wondered if some of the ways in which we routinely use human systems to improve quality and accuracy of information (e.g. use multiple sources) could be used similarly with AI systems. MP mentioned that in discussions with Alan Simpson[2] , Alan had mentioned the idea of an AI checker '2nd opinion' so a separate system which has the function of checking general AI system outputs.

MP also showed some slides with Alan's initial thoughts on areas of applicability, issues and risks with AI systems, highlighting where data is involved. Some of these are:

1. Safety related systems / services
2. Safety related procedures / human operations
3. Development of systems / services
4. Safety assessment and assurance

E.g.

a. Use of AI to support the implementation of operational procedures
   Can be used for process/procedure creation, analysis and review
b. Use of AI in the design of safety related solutions
   e.g. code generation, verification, validation, etc.
c. Use of AI to supplement or enhance safety assessment and analysis processes
   e.g. HAZOPs, FTAs, FMECA, etc.
d. Use of AI to support language aspects of safety work

---

[1] DB only tried ChatGPT during a lunchtime discussion and did not spend more than a few moments to consider in depth the requirements text that it produced; for example to look for inconsistencies.

[2] Alan Simpson will head the new SCSC Safe AI Working Group – SAIWG to be launched at SSS'24, https://scsc.uk/e1007

e.g. writing safety reports, safety cases, safety related research, etc.

e. Use of AI to support the review of documents
e.g. next level spelling & grammar, inc. detection of fallacious arguments
Review use of language, such as conciseness, technical correctness, relevance and audience understanding.

Also some Technical Reliability and Performance areas:

- **Accuracy**: Ensuring AI systems make correct decisions based on the data provided.
- **Robustness**: The system's ability to handle unexpected inputs or changes in the environment without failure.
- **Explainability**: The decisions made, especially those that impact safety, should be understandable to human operators.

And Data and Privacy Concerns:

- **Data Security**: Protecting safety-critical data from unauthorized access and ensuring integrity.
- **Privacy**: Ensuring personal data is handled in compliance with privacy laws and ethical standards.
- **Bias and Fairness**: Mitigating biases in training data which can lead to unfair or unsafe outcomes.

It was suggested that tools like ChatGPT could make people incredibly lazy, at least they would not bother to do the initial research or check sources.

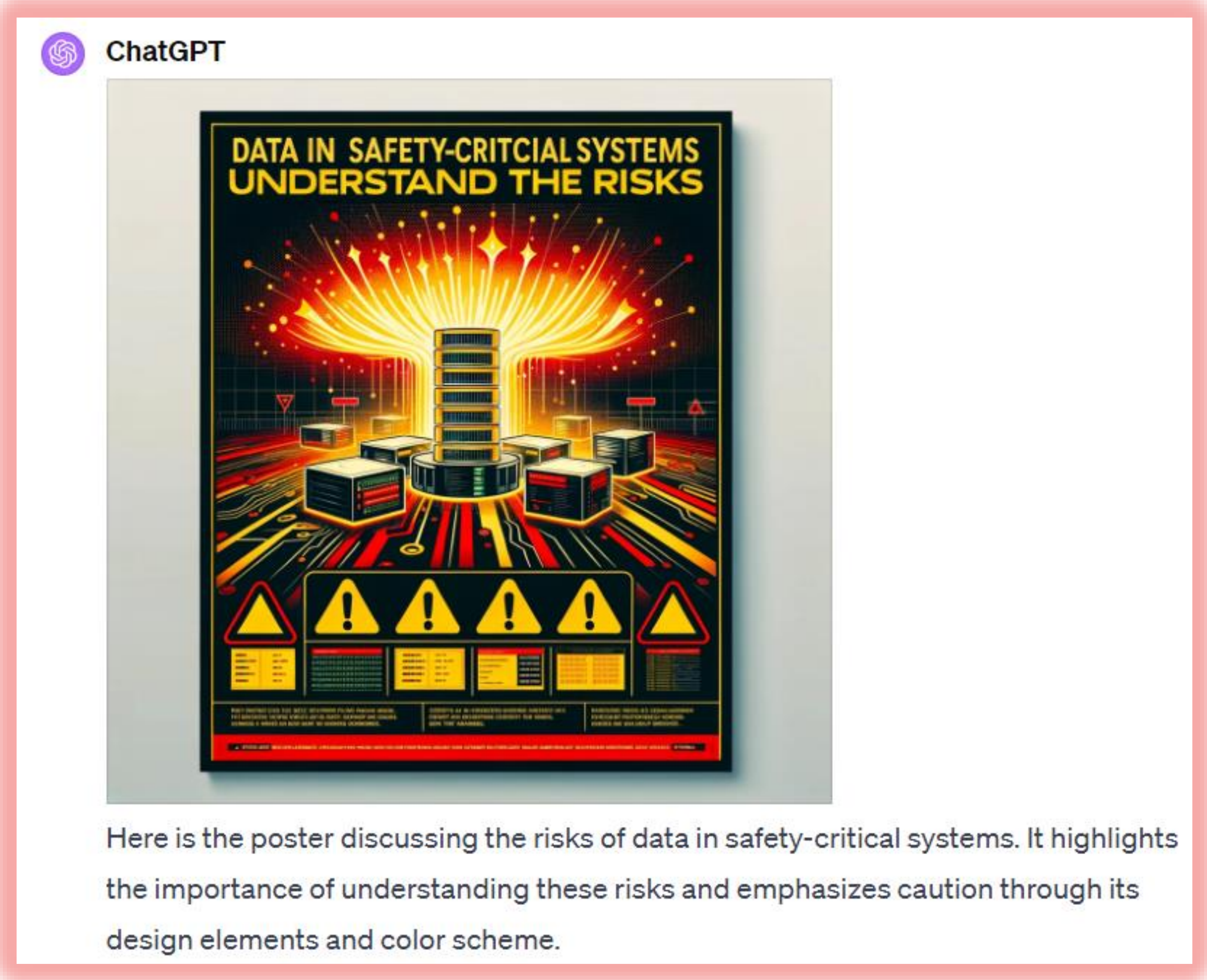
3. Poster for SSS'24

The DSIWG poster for SSS'24 in February next year, https://scsc.uk/e1007 was discussed.

TomTom were unable to help with the update of the poster (it was thought use of some TomTom map data as a central image in the poster might have been good).

Volunteers to help with the poster update were requested. It was suggested that DALL-E might be able to help(!).

[MP had a go with DALL-E after the meeting and it came up the following:



ChatGPT

Here is the poster discussing the risks of data in safety-critical systems. It highlights the importance of understanding these risks and emphasizes caution through its design elements and color scheme.

This image looks quite attractive and has some good elements. Unfortunately the text is too small or is meaningless at the bottom of the poster, however if this was replaced with the text with the earlier conversation, the result would probably be a good start.[3]]

ACTION 81.1 (MP) – Draft update to DSIWG poster for SSS'24

4. New Guidance Version

Updates which are likely to make it into the next version of the guidance document to be issued at SSS'24 were discussed:

1. Corrections and minor wording updates
2. Reworking of the accident list - new candidates in footnote:[4]

---

[3] DB suggested that one consideration for the group is the possible contradiction of warning against the use of AI whilst at the same time leveraging it. May be in that sense the "skilled" use of an AI to "assist" in a complex task could be highlighted? E.g., "Image generated by DALL-E, words by humans."

[4] [AM]: Article about Cruise dragging pedestrian: https://techcrunch.com/2023/11/08/cruise-recalls-entire-fleet-after-robotaxi-ran-over-dragged-pedestrian/; Cruise released data on pedestrian drag: https://getcruise.com/news/blog/2023/important-updates-from-cruise/; Fines and demands for evidence:

3. Work some appendices into the body of the document
4. New appendix on tooling, mentioning that ‘Radish’ is available. Possibly also showing an example of application

There was a general discussion about the impact of AI and moral and ethical issues. False information (‘hallucinations’) was thought to be a big issue.

The guidance document update editorial team comprises MT, TR, PL and MP. The update was planned for end of December. [Note this has not happened, but there is a meeting 03/01/24. Contact MP for details.]

NH suggested we may need bigger ‘laws’ governing behaviours of safety-critical systems. He gave an example of a pepper sorting machine which killed a man,
https://www.theregister.com/2023/11/09/robot_kills_employee/
https://www.bbc.co.uk/news/world-asia-67354709

5. DSITN (Data Safety In The News)

There was a possible data-related incident at Sellafield related to hacking of systems:
https://www.theguardian.com/business/2023/dec/04/sellafield-nuclear-site-hacked-groups-russia-china

6. Update on Tooling

DA and MA provided an update on the RADISH tool. There was currently a problem where new projects could not be created. This will be investigated.

7. Actions

See table at end.

8. AOB

ISO 26262: DB mentioned the update of ISO 26262. Edition 3 will have to include much more about data. RR said that Editions 1 & 2 assume that vehicles don’t communicate with each other, but Edition 3 will have to include data streaming in and out of vehicles. Data sources outside the vehicle are many and are potentially very complex. [Presumably the vehicle will have to protect itself from bad data generated from elsewhere.]

Deep Data: RR mentioned in the chat: Is there a possibility to split the two types of "Dark Data" described in the book along the lines of the descriptions of the "Deep Web" and the "Dark Web":
https://www.britannica.com/story/whats-the-difference-between-the-deep-web-and-the-dark-web
So the "Deep Web" could correlate to the "Known Unknowns" - so could be "Deep Data" and the "Dark Web" could correlate to the "Unknown Unknowns" and that is then the truly "Dark Data".

---

https://techcrunch.com/2023/12/04/cruise-faces-fines-in-california-for-withholding-key-details-in-robotaxi-accident/
[RR]: https://priuschat.com/threads/tesla-runs-red-light-into-a-semi-trailer-truck.223898/

9. Next Meeting

The SCSC Data Safety Initiative Working Group meeting #82 (DSIWG#82) will be held 11th January 2024 by Zoom:
https://us02web.zoom.us/j/89921972252?pwd=RHdhTTR6N1FiV3VuL3JIaUxDdWt5Zz09

10. Thanks

Thanks to all who provided contributions.
Thanks to MP for chairing and taking minutes.

Summary of Open Actions

Actions greyed out are considered closed and will be removed from the list at next issue.

Ref	Owner	Description	Target Guidance Version
42.9	MP	Work out a matrix of data categories (previously 'types') and data properties (as per DB discussion)	N/A
43.4	MP	Write up a data focussed FMEA approach.	4.0
44.2	MP	To discuss with AK on how to get the Wikipedia article published	N/A
46.1	MP	Review the application of DSALs to higher level forms of aggregation	N/A
53.1	MP	To talk to Kevin King about what we need to do in the guidance for digital twins.	4.0
61.2	AW	Research the relevance of digital currencies and report back to the group (with MA and MT)	N/A
63.1	CT	Look at both Dark Data and Dazzle Data for sensors (e.g. when a sensor is saturated, in noisy environment or when readings are below the detection level floor)	4.0
69.3	PMcK	Develop a scoping diagram that shows how the DSG fits into the overall lifecycle process and other standards	4.0
69.4	MA	Write a short note on the issues of aggregation	4.0
69.6	MA/DA	Update the data safety tool to use the latest version of the guidance document	-
71.3	PH/DA/RO	Develop security properties thinking further for next DSIWG	4.0
71.4	PH/DA/RO	Present security properties work to next SISWG meeting	-
71.5	AM	(i) Establish if any of this can be published within the DSIWG and (ii) Consider a structuring similar to that used in security standards or ISO26262	-
71.7	MP/CT	Consider impact of FAIR data on the guidance	4.0
73.1	JK, RR	Consider production of a short note which could be used as an appendix to the guidance on lessons learnt using the guidance at TomTom	4.0
73.2	MT	Consider how the guidance fits with different lifecycles considering 'V', Continuous Service, Agile and 'Data Pipeline'	4.0
73.4	DA	See whether any of the DSITN entries from previous minutes and slides could be used to enhance the list of accidents	-
75.3	MT	Fix the minor typos that have been reported with v3.5 and prepare updates for both the online version and also the KDP/Amazon hardcopy.	3.5 Update
76.1	MP, PH, MT, TR	Consider whether Guidance Annexes can be integrated into the main body of the document.	4.0
76.2	TR	Look further at IEC 25012 to consider whether a new Annex to the Guidance would be beneficial.	4.0

Ref	Owner	Description	Target Guidance Version
76.3	MT	Provide slides from last October's SQEPtember presentation (an introduction to data safety) to MP.	-
78.1	MP/MA	Investigate representation / format issues further. Consider adding another issue to the next version of the guidance on representation.	4.0
78.2	MP	Investigate data-data interactions further.	4.0
79.1	MT	Introduce EBA / EBTA with a short presentation at next meeting	
79.3	PH	Report on use of the RADISH tool at next meeting	
79.4	PG	Re-visit the earlier nuclear data safety work using the RADISH tool	
80.2	TR	See which appendices are candidates to work into the main body of the document	4.0
80.3	DA, MA	Write appendix on tooling and RADISH	4.0
80.4	DA, MA	Update the data-safety.tech web site with all known data safety accidents (from past DSIWG minutes and any other sources).	
80.5	MT	(i) Add a ref to the data-safety.tech website in the guidance and (ii) Use the updated data-safety.tech website to produce a better spread of accidents for the new version of the guidance	4.0
81.1	MP	Draft update to DSIWG poster for SSS'24	

Additional chat:

16:11:59 From Brent K To Everyone:
Keep in mind that the deep computing in memory AI hardware could be used for other purposes. ;)
16:13:07 From Brent K To Everyone:
https://www.jstage.jst.go.jp/article/transfun/advpub/0/advpub_2022CIP0023/_pdf
16:23:08 From Brent K To Everyone:
RE AI systems - How do we do full (and accelerated) lifecycle qualification, testing & validation... How do we know when the AI system needs to be requalified, tested, and validated?
16:40:12 From Brent Kimberley To Everyone:
My apologies for the link. The link pointed a paper entitled "Multiparallel MMT : Faster ISD Algorithm Solving High-Dimensional Syndrome Decoding Problem" which relied on 10 "AI machines" GPU machines. It was published by Shintaro NARISADA, Kazuhide FUKUSHIMAand Shinsaku KIYOMOTO on or around 2022/11/09 in the IEICE TRANS. FUNDAMENTALS, VOL.E106–A, NO.1. JANUARY 2023.
17:00:11 From Brent Kimberley To Everyone:
As an aside, we struggle to fuse simple absolute values like an odometer or run hours various systems of records. Maintenance Depots (aka mechanics) are often 1 sigma with respect to time or value (it's the weekend, let's leave the paperwork until Monday). Operators sometimes enter one too many or one too few digits. And, automated systems are often 3 sigma or more. And accountants think everything should be accurate to the penny.
17:00:37 From Brent Kimberley To Everyone:
run hours from our various (component) systems of record.