ISA Working Group 2020 Webinar Series

(Webinar will commence at 11:02 am)

Are you interested in joining the ISA Working Group? Let us know by e-mailing SEP@theiet.org

Assurance in a Connected World

Webinar 5: Functional Safety and AI

Panellists

Audrey Canning – Speaker

Stephen Hatton – ISA WG Chair

Pete Hutchison – ISA WG Deputy Chair

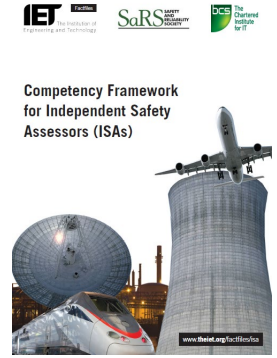
John Canning – ISA WG Secretary

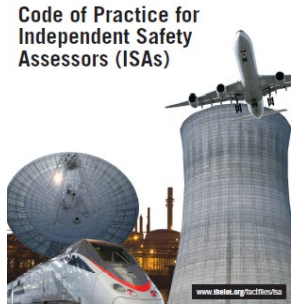
ISA WG Terms of Reference – Purpose

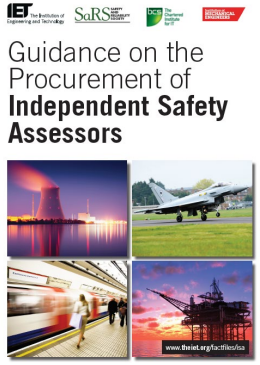
- Promote the ISA role as a means of providing independent safety assurance of products to the supplier, purchaser and user
- Promote the ISA role of a safety professional in standards
- Support professional development by defining minimum standards, identifying training that meets minimum standards and supporting resources
- Support professional ISAs by developing guidance and providing information that affects their role

Guidance – Published

- General
  - ISA Working Group Terms of Reference
  - What is Independent Safety Assessment (ISA)? in Review for update
- Professional
  - ISA Code of Practice for Independent Safety Assessors (ISAs)
  - Competency Framework for Independent Safety Assessors (ISAs)
- Substantive Guidance
  - Assessment of Safety Related Compliance Claims (SRCC)
  - Guidance on the Procurement of Independent Safety Assessors
- Guidance Notes / Position Papers
  - Guidance on the Use of Accident and Incident Data by ISAs
  - Documents useful to Independent Safety Assurance
  - Position Statement on Security, Safety and ISA

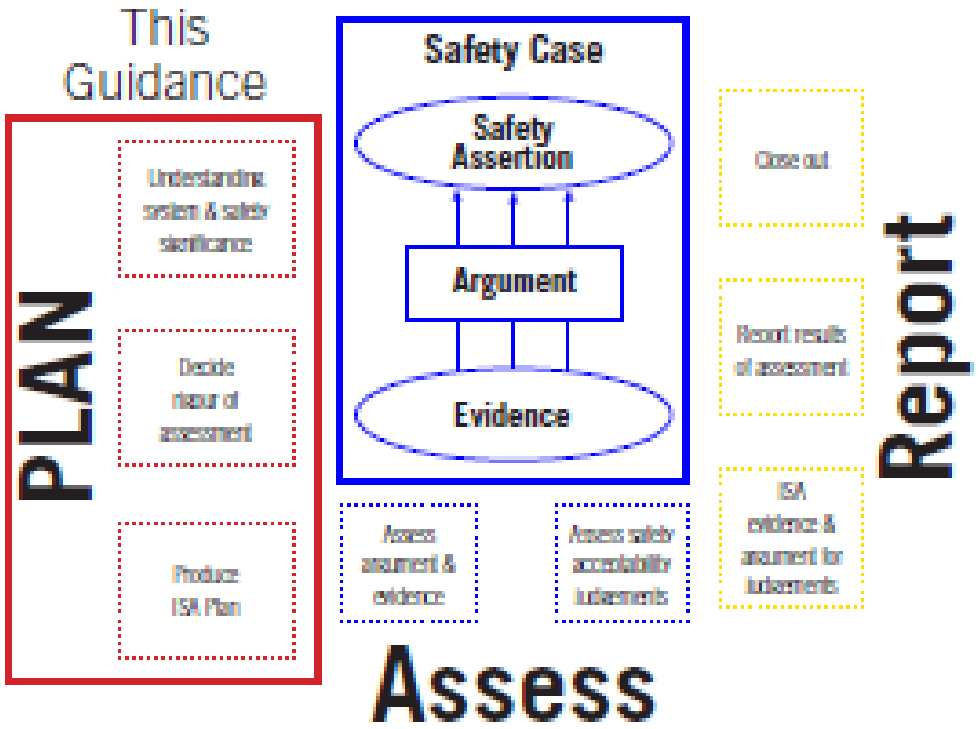






Assessing a Safety Case Series

- Assessing a Safety Case Series
  - Guidance for Producing an ISA Plan for Assessing a Safety Case — published
  - Guidance on Safety Assessment Reports — to be published
  - Guidance on Degree of Rigour — in progress



Documents in Development

- Standards Group
  - Requirements for independent review/assessment called up in Standards and Industry Guidance
  - Environment Assurance and Safety Assurance

- Professionalism Group
  - Using Key Performance Indicators with an ISA Contract ready for issue
  - Agile Development

Housekeeping

- Q&A (Zoom Webinar)
  - Use Q&A button to type your question (don’t use chat button; don’t raise hand)
  - Use ‘thumbs up’ to vote up or vote down a question (once only)
  - Panellists will select and pose questions on your behalf
  - Questions not discussed today will be recorded and commentary provided afterwards
- Feedback
  - Short re-cap article after the event
  - Please read and complete our questionnaire (to be e-mailed to you)
    - What are your thoughts on functional safety and AI?
    - No need to answer all questions
  - Let us know if you’re interested in joining the ISA Working Group

Functional Safety and AI

Audrey Canning

Audrey has more than 35 years of experience in Functional Safety (and before that six years experience in the of development of digital systems and two years experience in the development of AI based systems). She is currently the Convener of the software engineering aspects of IEC 61508, the world-wide representative for functional safety on the IEC ACOS committee reporting to the IEC Standards Management Board and the IEC/SC65A liaison member to the Joint Working Group between IEC and ISO, JTG1/SC42.

She is also the Middle Warden of the Worshipful Company of Engineers, a Fellow of the IET and a member of the IET Engineering Safety Panel.

Functional Safety and AI

Audrey Canning
Virkonnen Ltd
Convener IEC 61508-3 Maintenance Team
IEC SC65A Liaison Member to ISO/IEC JTC1 SC42

December 2020

Virkonnen

Introduction

- Functional Safety Requirements vs Ground Reality
- A Joint ISO/IEC Initiative
- Frameworks and Language
- Working it Out
- Concluding Remarks

FS Requirements vs Ground Reality

- BS EN 61508 (last published June 2010)

Table A.2 – Software design and development – software architecture design

(see 7.4.3)

	Technique/Measure *	Ref.	SIL 1	SIL 2	SIL 3	SIL 4
4b	Graceful degradation	C.3.8	R	R	HR	HR
5	Artificial intelligence - fault correction	C.3.9	---	NR	NR	NR
6	Dynamic reconfiguration	C.3.10	---	NR	NR	NR
7	Modular approach	Table B.9	HR	HR	HR	HR

---	the technique or measure has no recommendation for or against being used.
NR	the technique or measure is positively not recommended for this safety integrity level. If this technique or measure is used then the rationale behind using it should be detailed with reference to Annex C during the safety planning and agreed with the assessor.

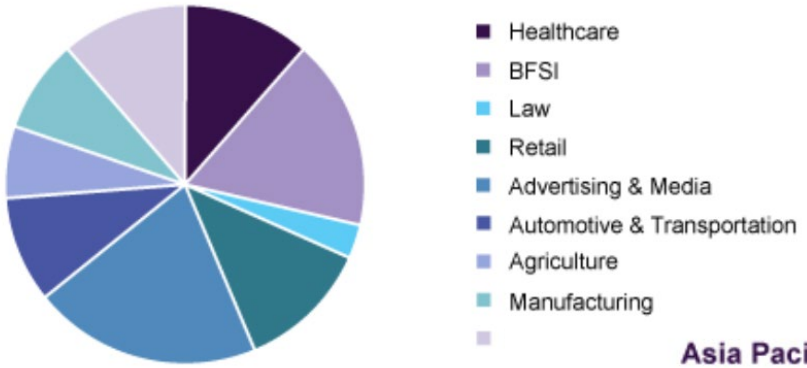
Virkonnen

FS Requirements vs Ground Reality

- Financial Analysis
- Social Media / Chat bot
- Face Detection and Recognition
- Disease mapping / Proactive healthcare management
- Digital Assistants
- Manufacturing robots
- Critical machine health monitoring
- Autonomous vehicles

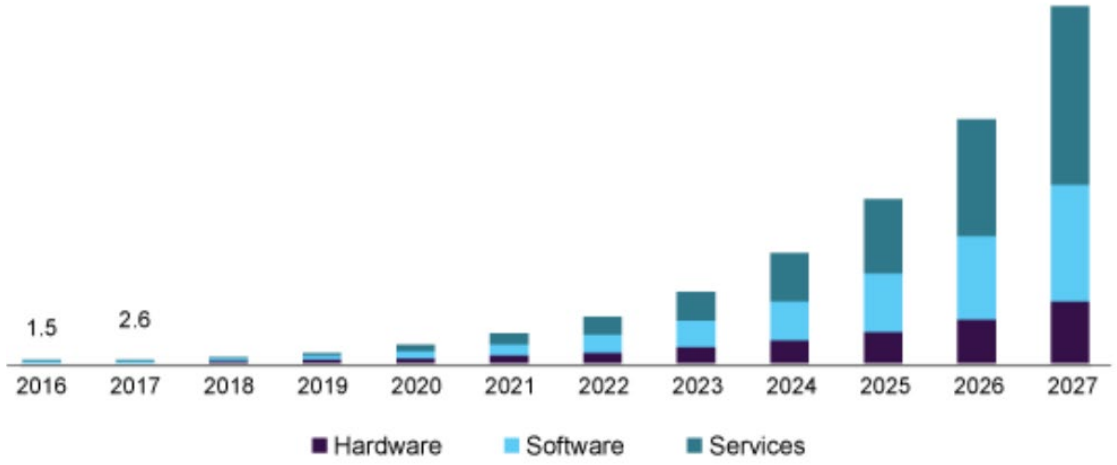
FS Requirements vs Ground Reality

Global artificial intelligence market share, by end use, 2019 (%)



Source :
https://www.grandviewresearch.com

Asia Pacific artificial intelligence market size, by solution, 2016 - 2027 (USD Billion)



FS Requirements vs Ground Reality

Challenges

- Unproven ‘physics‘/ extrapolation difficult
- Incomplete/biased data
- Human interpretation
- False re-enforcement
- Behaviour changed from validated system
- Could transfer ethical decisions to machine

A Joint ISO/IEC Initiative

- 9/18 – agreed to discuss future strategy/positioning of MT-3 with respect to new computational technologies
- 7/19 – Agreed:
  - At a minimum clarify whether AI/autonomy is banned
  - If not, prepare guidance
- 9/19 – approached ISO/IEC JTC1 SC42 (working on AI) to propose joint TS on Functional Safety and AI
- 12/19 – ISO/IEC JTC1 SC42/WG 3 member attended MT61508-3
- 1/20 – MT61508-3 Convener attended ISO/IEC JTC1 SC42/WG3
- 3/20 – Joint New Work Item Proposal for a Technical Report prepared

A Joint ISO/IEC Initiative

- TR Objective: to describe the properties, related risk factors, available methods and processes relating to use of AI :

- Inside a safety related function to realise the functionality
- To control equipment, but protected by non-AI based safety related functions to ensure safety
- In toolchains used to design and develop safety related functions.

- 4/20 – Joint voting on NWIP (a Technical report) in ISO & IEC
- 5/20 – NWI approved, commenced Task Group

Frameworks and Language

- Risk
  - ISO - the "effect of uncertainty on objectives" and an effect is a positive or negative deviation from what is expected.
  - IEC - combination of the probability of occurrence of harm and the severity of that harm
- Harm – physical injury or damage to the health of people, damage to property or the environment
- Functional Safety - part of the overall safety relating to the EUC and the EUC control system that depends on the correct functioning of the E/E/PE safety-related systems and other risk reduction measures

Frameworks and Language

AI Technology Class / Usage Level	Techniques meeting current 'safety properties'	Techniques with some shortfall in meeting current safety properties, but mitigations can be identified	Techniques unable to meet current safety properties and where suitable mitigation cannot be identified - safety has to be assured external to the AI
Directly in safety loop			
Indirectly affects safety loop – e.g. diagnostics			
Used during development– with decision making			
Used during development – no decision making			
Non-safety system, but places demand			
System that can be shown through HA not to affect safety			

Working it Out

Conducted risk assessment per ‘cell’ to understand issues

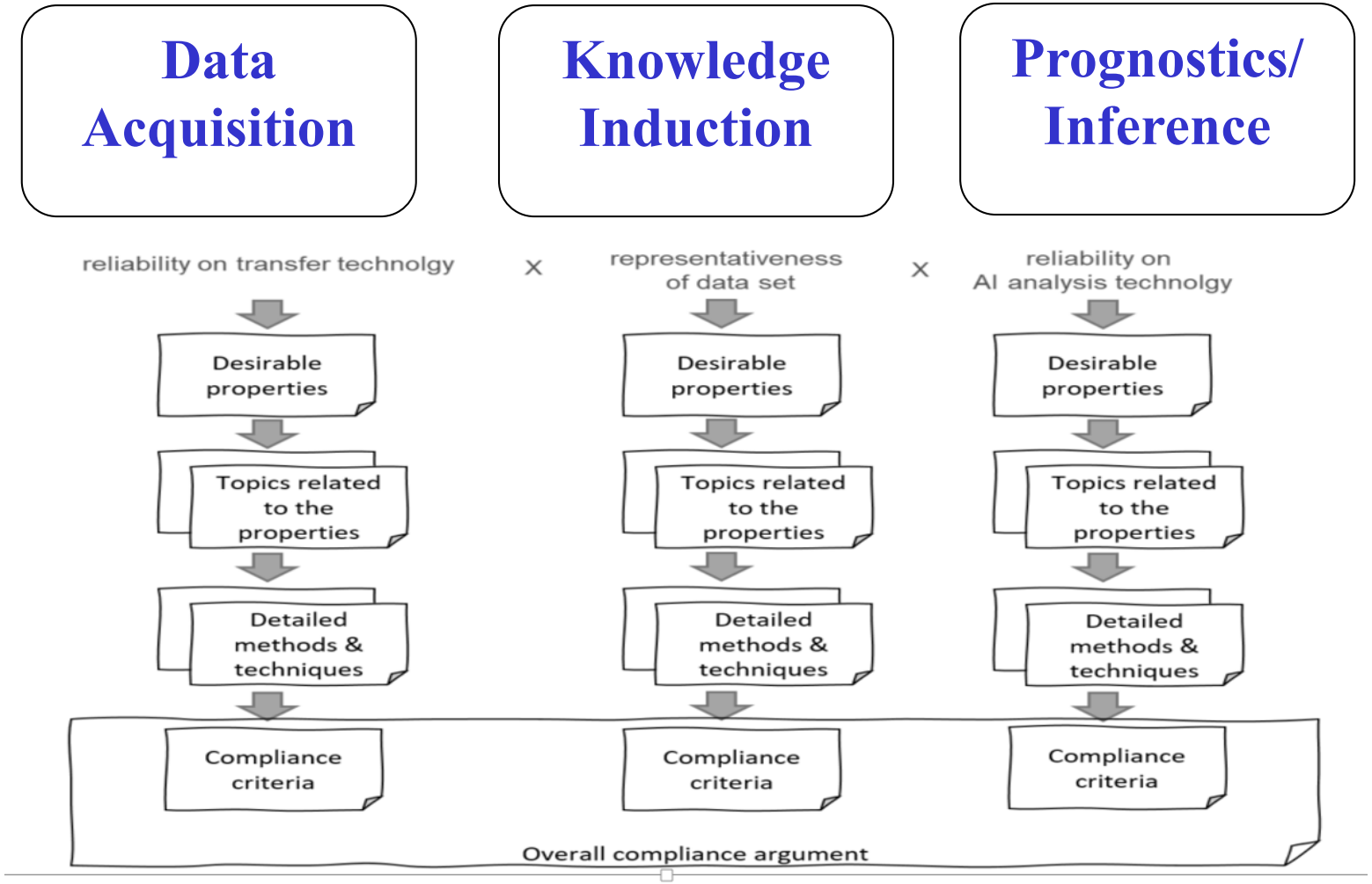
Class	Use	Technology?	What are the local consequences on 61508 properties?	How can we mitigate?
Techniques able to meet the ‘properties’ underlying existing safety standards	A1 AI technique used in a safety relevant E/E/PE system and automated decision making e.g. Automated breaking system with pedestrian recognition of an autonomous vehicle.	Decision tree Programmed algorithm Rule based system Predicate logic Parameterisation Switching systems Database DNN which is sufficiently simple that it could satisfy Class 1. -	Difficult to demonstrate completeness, correctness, understandability. Likely to have intrinsic specification faults or ambiguity Difficult to demonstrate freedom from adverse interference of non-safety functions with the safety needs Difficult to clearly identify verification and validation methods	Identify IEC 61508-3 techniques (or techniques through 61508-3 Annex C with equivalent objectives with respect to the lifecycle) that achieve all the six desirable properties Examples: use of ‘above normal’ rigorous V&V activities to show properties achieved.
Techniques unable to meet safety properties and where no suitable mitigation can be identified	A1 AI technique used in a safety relevant E/E/PE system and automated decision making e.g. Automated braking system with pedestrian recognition of an autonomous vehicle.	Mathematical non-linear function where the number of parameters adequate to represent system is unknown (e.g. neural network - NN) Deep learning – e.g. multi-layer NN Base data which is not repeatable – e.g. an uncontrolled environmental image vs a predefined image System which is not ‘bounded’ – e.g. operating outside known behaviour, systems of systems with emergent properties	Intended output is not predefined – environment is not bounded Existence of ‘tipping points’ not recognised in represented by the algorithm. The algorithm may only be able to identify a sub-set of possible solutions (e.g. cats and dogs analogy)	Bounding the parameters – e.g. predicting the p(failure) for Kalman filters. Dynamic risk management (research) – measure risk of harm during operation - or add layer of protection – introduce safety bag/cage (e.g. separate emergency brake from control)

nen

Working it Out

- Conducted case study ‘trial’ for three applications:
- Automotive example - AI directly in safety loop and not easy to demonstrate against conventional standards
- Robot safety protection function - speed and separation between human and machine - directly in the safety loop, partially justifiable against conventional standards
- Oil & Gas machine health monitoring, providing data to human decision makers, partially justifiable against conventional standards
- Currently in the process of extracting the main ‘process steps’ the ‘common features’ to populate a framework with the goal of determining approaches to demonstrate ‘an equivalent level of safety’

AI Functions Workflow



Virkonnen

Other Issues

- The ‘boundedness’ of the input data (and the simplicity of the sensor inputs) has an impact on the ability to validate an AI system functions (e.g. a motorway vs and unconstrained town scenario); the ‘framework’ needs to be extended to address this aspect

Virkonnen

Other Issues

- The ‘boundedness’ of the input data (and the simplicity of the sensor inputs) has an impact on the ability to validate an AI system functions (e.g. a motorway vs and unconstrained town scenario); the ‘framework’ need to be extended to address this aspect
- It is likely that different ‘risk mitigation methods’ will be needed for different types of AI techniques
  - at the ‘methods level’
  - at the ‘systems level’
- ‘SIL’ and ‘Systematic Capability’ not yet part of framework

Concluding Remarks

- Not an easy topic:
  - 2 different cultures /languages /drivers /knowledge
  - many AI technologies
  - insufficient examples of ‘success’ for a ‘safety recipe’

- Continuing to analyse the results of our case study to extract guidance on:
  - completeness of the safety properties for different stages
  - hierarchy of functional ‘elements’ with an AI implementation
  - potential system and technology level mitigations

Concluding Remarks

- Some (safety practitioner) consensus that:
  - Complex AI systems not appropriate within the safety loop
  - but could be appropriate for some ‘indirect’ safety applications subject to a hazard and risk assessment and identification of appropriate mitigations, both system and technique level
- On-line learning’ – difficult to show does not invalidate ‘V&V’, unless system level mitigations could constrain ‘learnt’ behaviour within safe bounds

- But it does look like it is possible to prepare a framework and route map for the type of rationale that would be needed if one intends to use MT61508-3 NR techniques.

Virkonnen

Concluding Remarks

‘The main reason the 1956…..workshop didn’t live up to my expectations is that A.I. is harder than we thought’.
John McCarthy 2006

Whilst the views expressed in this presentation are those of the author, she is in debt to the ideas and discussions of the SC65A Liaison Group with ISO/IEC JTC1 SC42, especially Takashi Egawa the Liaison Member from WG3

AT E Schoitsch K Meyer-Gräfe **DE** C Gregorio F Poignee H Laible	**DE** M Kindermann M Kollmann P Feth R Adler S Aschenbrenner T Boemer T Loeffler	**FR** B Ricque S Dissoubray **IR** T Meany **IT** R Mariani	**JP** H Kanamaru N Kanekawa Y Cheng Y Oiwa Y Sato **NO** T Myklebust	**UK** A Canning D Daniels R Morgan **USA** A Mishra S Visalli

Virkonnen

Thank You for Attending

Please join us again for our next seminar or series of webinars

Date(s) to be confirmed

Are you interested in joining the ISA Working Group?

Let us know by e-mailing SEP@theiet.org