UK Atomic Energy Authority

Functional Safety and Artificial Intelligence

Christopher Hume (UK Atomic Energy Authority)

20th Nov 2024

RACE

Official

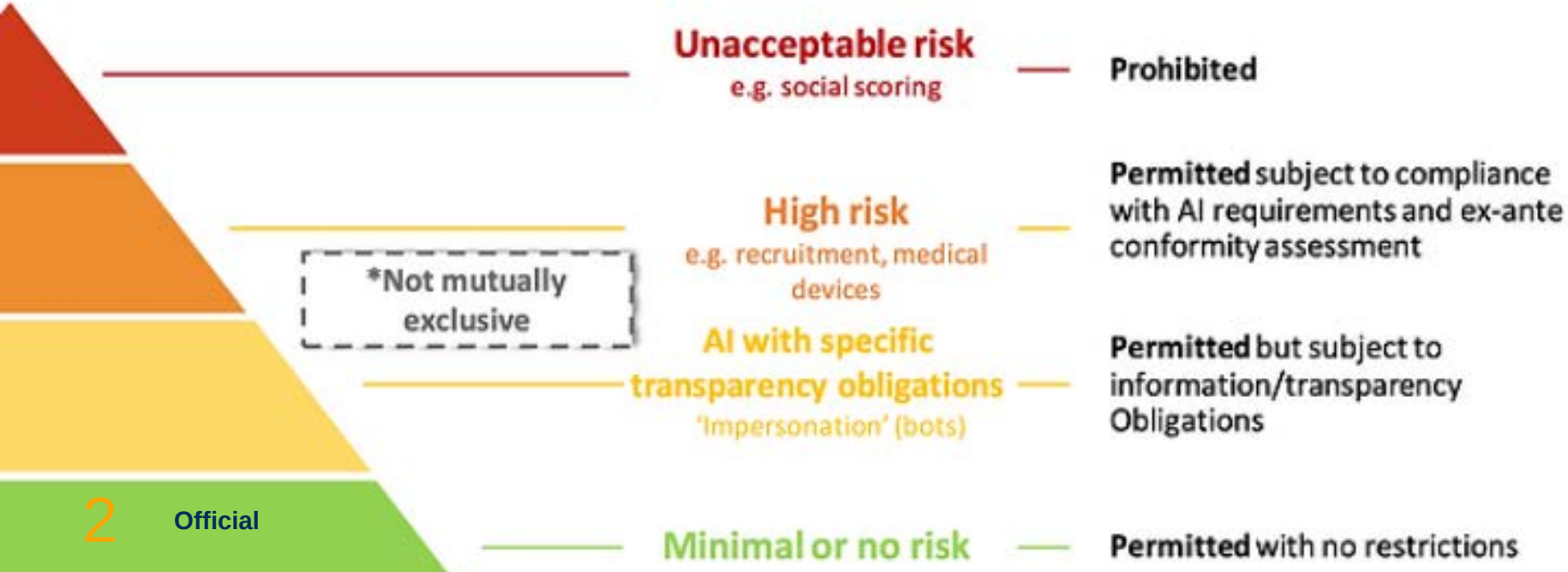
Regulation

UK AI Regulation Bill

- AI Authority
- AI Responsible Officers
- Regulatory Sandboxes
- Records Transparency

EU AI Act

- Within a four-tier system any use of AI in Safety Components or in Critical Infrastructure is to be considered High Risk.
- For High-Risk systems developers must:
  - Ensure data quality
  - Systems capability assessment and limitations
  - Design to allow for effective Human Oversight



Standards

- ISO/IEC TR 5469 (22440): Artificial intelligence — Functional Safety and AI Systems.
  - Extends IEC 61508 – defines terminology and considerations but not rules to achieve integrity levels.

- VDE-AR-E 2842-61: Development and Trustworthiness of Autonomous Cognitive Systems
  - Extends IEC 61508 – addresses trustworthiness as an expansion of the safety concept to incorporate elements such as ethics

- UL4600: Standard for Safety for the Evaluation of Autonomous Products
  - Focused on Autonomous Vehicles and compatible with IEC 26262 and ISO/PAS 21448.

Figure 2: Aspects of Trustworthiness

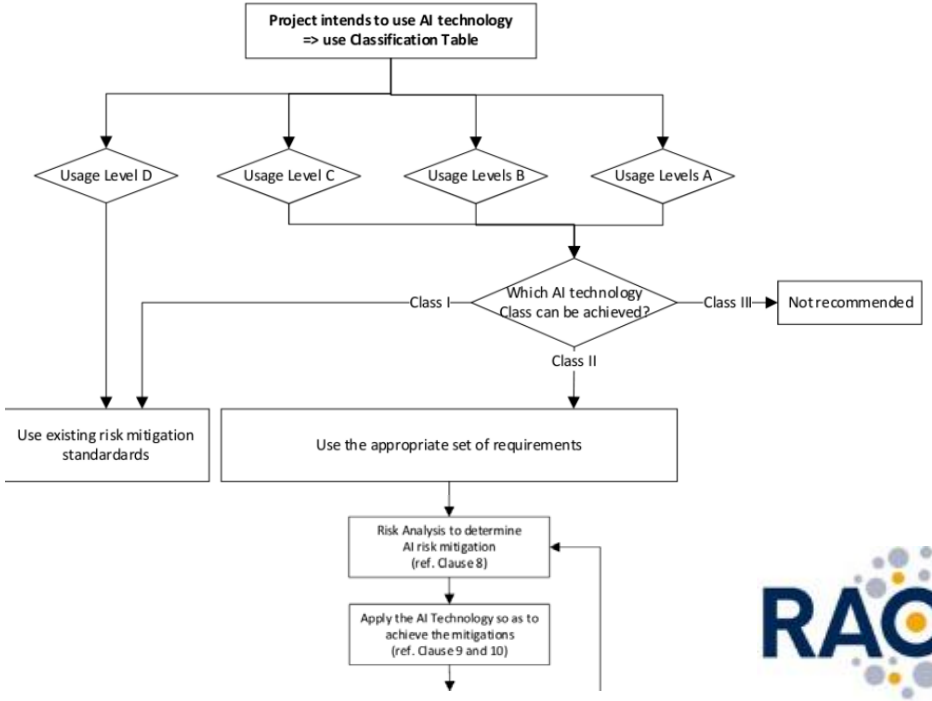
Classification and Evaluation

- Technology Class – The degree to which the AI technology can be developed and reviewed under existing Functional Safety Standards and frameworks.
- Usage Level – Whether the AI technology is used within:
  - Operation or Development
  - Primary Function or Diagnostics or Indirect impact (such as demand rate)

Advanced Technology Class and Usage Level should not be Combined!

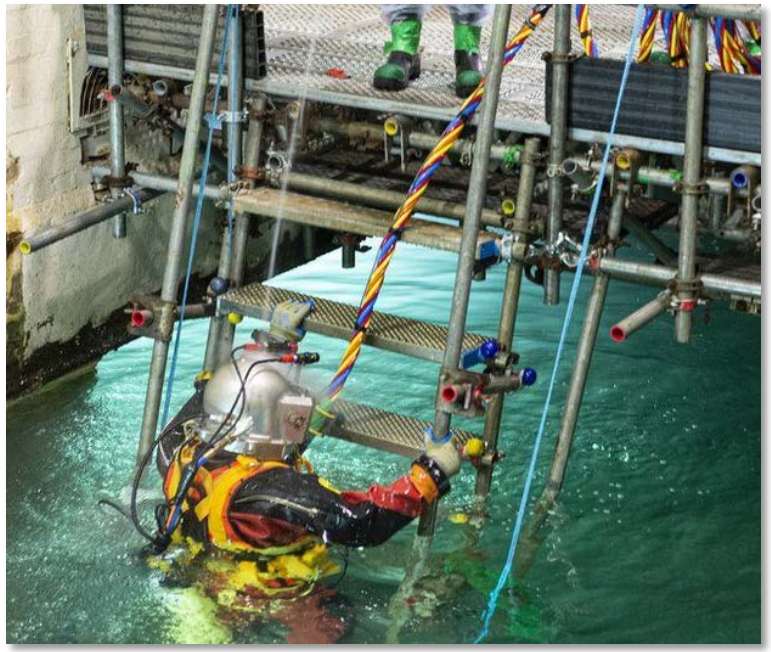
- Maximising Explainability and Transparency
  - Network Visualisation and Inspection
  - Input Masking

- Environmental Complexity
  - Appropriate Requirements and Reward Function



Nuclear Decommissioning – Use for AI

UK Atomic Energy Authority



Decommissioning with human operators will always put human operators at risk

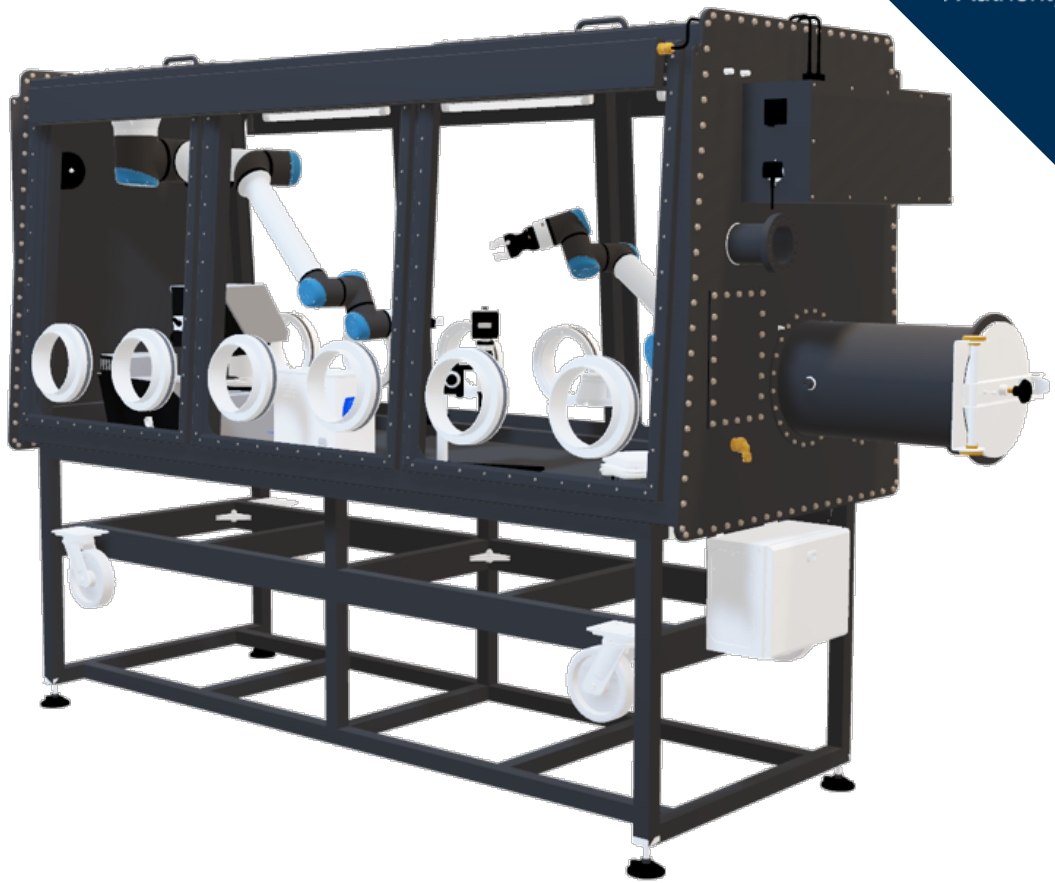
Lots of protective equipment, high supervision, highly procedural, slow

Difficult tasks to automate, complex environments,

RACE

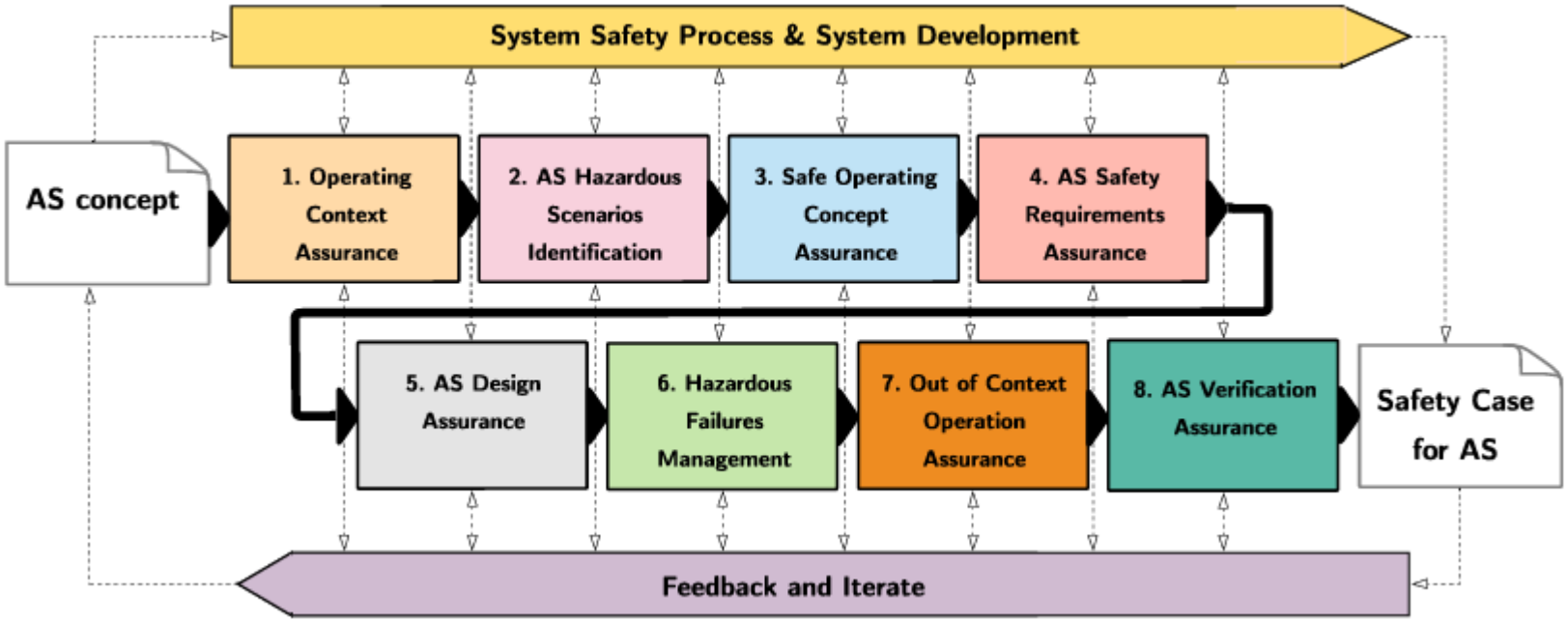
RAICo Glovebox

First of a kind Autonomous Glovebox
Safety Case Development

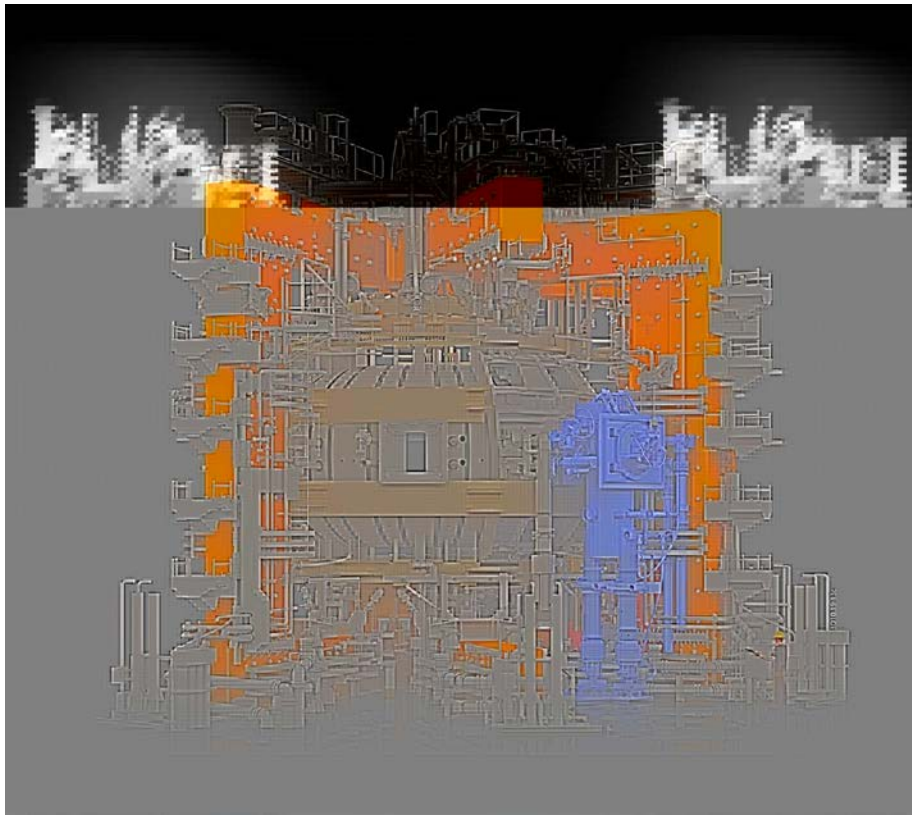


SACE

- Guidance on the Safety Assurance of Autonomous Systems in Complex Environments (SACE) by the Assuring Autonomy International Programme (AAIP) at the University of York.
- A safety framework utilising systems engineering principles to plan and carry out an AI safety case for complex systems.



Use case for AI in fusion



Fusion Tokamak

- Future global fusion opportunity
- Proven, but fusion at-scale inevitably requires enhanced use of technology
- Challenges involve:
  - Sense and control
  - Disruption mitigations
  - Big Data
  - Radiation (short-lived)
  - Complex System-of-Systems

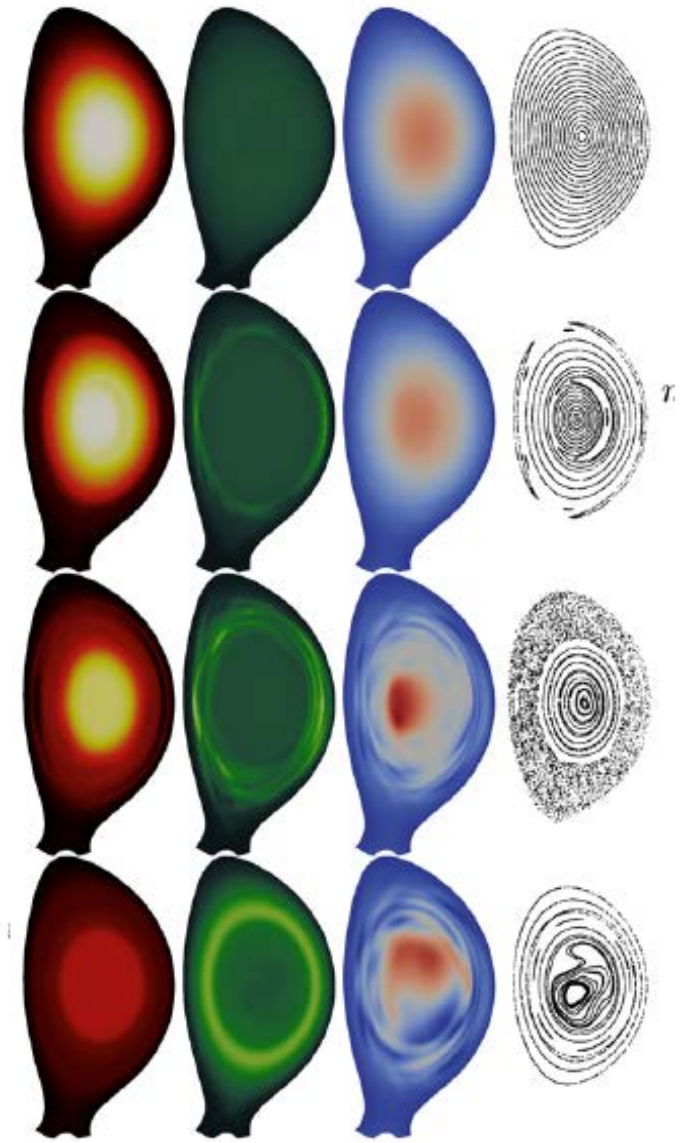
Plasma Instabilities As A Hazard

- Plasma instabilities threaten the viability of fusion reactors due to significant:
  - Reactor vessel damage
  - Damage to supporting systems and utilities particularly divertors
  - Plasma Loss
  - Some risk to humans but lower than Fission applications.
- Difficult problem for safety problem and control as instabilities grow rapidly and chaotically from microscopic turbulence.

AI Addressing Plasma Instabilities

- We have used Neural Networks to form surrogate models for the partial differential equations which predict plasma behaviour.

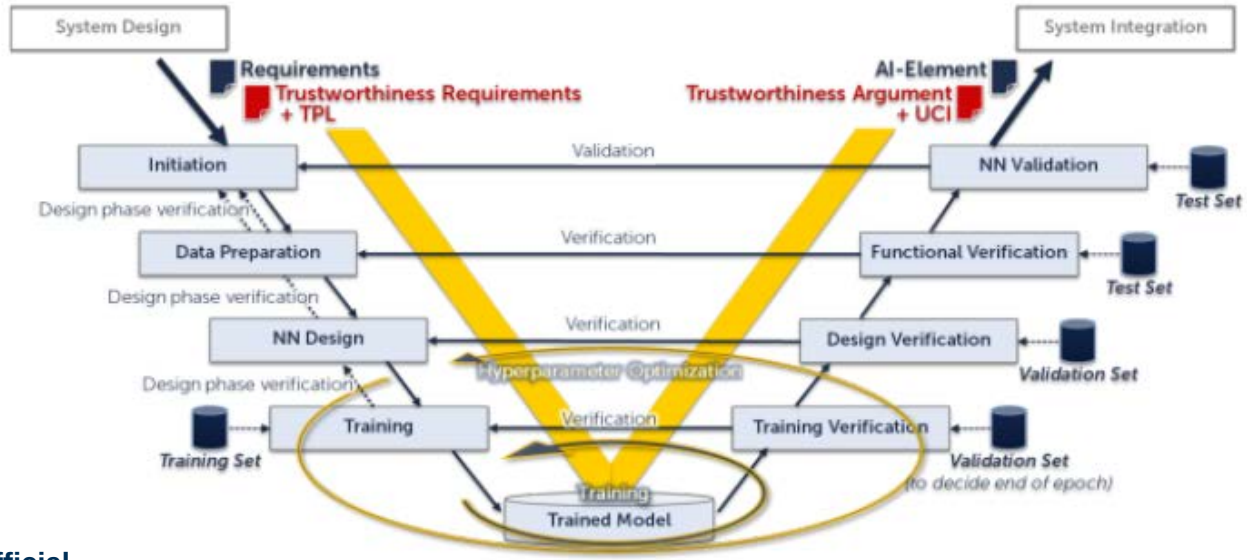
- Gaussian processes give Uncertainty Quantification for model accuracy and the action results.



Design

- Use of Supervising Functions
- Uncertainty related failures as a third failure type connected to forseeable misbehaviour
  - Uncertainty Confidence Indicators

type of failure	measures	HW measures	SW measures	AI measures
systematic	qualitative requirements	systematic capability	systematic capability	systematic capability
random	quantitative requirements	λ, SFF, DC, target values	-- / --	-- / --
uncertainty-related	structured approach	-- / --	-- / --	Uncertainty confidence indicator (UCI)



- Structured Approach
  - AI Blueprints & Design Patterns

Future of Plasma Control and Safety

- Virtual and Physical Testing (Shattered Pellet Injection)

Any Questions?

UK Atomic Energy Authority RACE

13:25	Slot A-7: ***Functional Safety and Artificial Intelligence (AI)*** Chris Hume – UKAEA RACE Jon Wiggins – ABD Solutions	Slot B-7: ***CASS 61511 Workshop*** Andrew Derbyshire – ERM Deepti Chauhan – Sensia
14:00	Slot A-8: ***Black Box Testing for Functional Safety*** Dr. Silke Kuball – EDF Energy	
14:35	Slot A-9: ***Machinery Functional Safety with IEC 62061 and ISO 13849*** Paul Reeve – SILMETRIC	Slot B-9: ***Functional Safety Tool Qualification*** Hassan El Sayed – UL Solutions