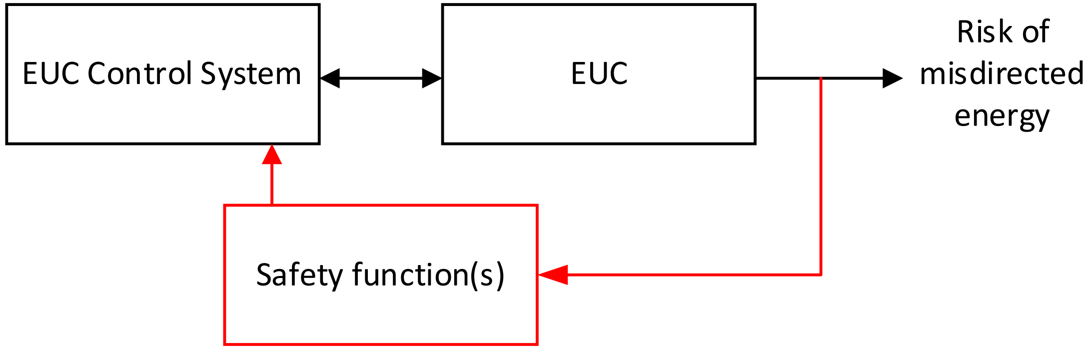
Functional Safety and AI technologies: background, standardization landscape and overview of ISO/TR 5469

Takashi Egawa, Riccardo Mariani

2nd ISO/IEC AI workshop, November 30th, 2022

Functional Safety



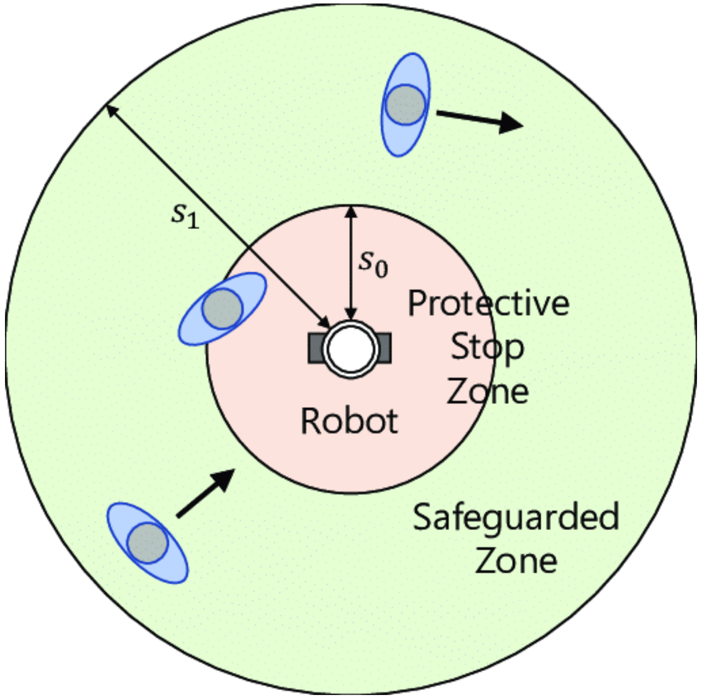
IEC 61508-4 defines functional safety as that "*part of the overall safety relating to the EUC (Equipment Under Control) and the EUC control system that depends on the correct functioning of the E/E/PE (Electrical/Electronic/Programmable Electronic) safety-related systems and other risk reduction measures.*"

The E/E/PE safety-related system is delivering a "safety function", which is defined in IEC 61508-4 as a "*function to be implemented by an E/E/PE safety-related system or other risk reduction measures, that is intended to achieve or maintain a safe state for the EUC, in respect of a specific hazardous event.*"

Example of safety functions

Emergency Stop

Protective Stop



Functional Safety Basics

- **Fault:** abnormal condition that may cause a reduction in, or loss of, the capability of a functional unit to perform a required function
- **Error:** discrepancy between a computed, observed or measured value or condition and the true, specified or theoretically correct value or condition
- **Failure:** termination of the ability of a functional unit to provide a required function or operation of a functional unit in any way other than as required

Systematic Fault (HW and SW)

- Human-induced, such as software bugs or hardware design marginalities
- Deterministic occurrence
- Focus is mainly on fault avoidance
- Mitigations primarily focus on process related improvements and diversity.

Random HW Fault

- Not human-induced, such as electromigration, radiation effects, interferences
- Non-deterministic: following a probability distribution
- Focus is mainly on fault detection
- Mitigations primarily focus on safety mechanisms and redundancy concepts

IEC 61508

- Overview:
  - It is a basic functional safety standard applicable to all kinds of industry
  - It addresses system/hardware and software
  - It introduces a safety lifecycle
  - It uses a probabilistic failure approach

IEC 61508-1

Edition 2.0 2010-04

INTERNATIONAL STANDARD

NORME INTERNATIONALE

BASIC SAFETY PUBLICATION

PUBLICATION FONDAMENTALE DE SÉCURITÉ

Functional safety of electrical/electronic/programmable electronic safety-related systems –
Part 1: General requirements

Sécurité fonctionnelle des systèmes électriques/électroniques/électroniques programmables relatifs à la sécurité –
Partie 1: Exigences générales

Functional Safety and AI

- The use of artificial intelligence (AI) technology in industry has increased significantly in recent years and AI has been demonstrated to deliver benefit in several applications.
- However, there is limited guidance on specification, design and verification of functionally safe AI systems or on how to apply AI technology for functions that have safety-related effects.

Safety AI landscape (1/3)

- **ISO/IEC JTC 1/SC 42**
  - TR 5469 – “Artificial intelligence — Functional safety and AI systems”
- **ISO/TC 22/SC 32**
  - ISO 21448 – Annex D.2 – “Implications for Machine Learning”
  - ISO/TS 5083 - Annex B – “Safety and cybersecurity for AI – application to automated driving systems”
  - ISO/PAS 8800 - “Road Vehicles – Safety and Artificial Intelligence”

Safety AI landscape (2/3)

- **IEEE**
  - IEEE CS FSSC – Functional Safety Standards Committee
  - IEEE P2851 - Standard for functional safety data format for interoperability within the dependability lifecycle

Safety AI landscape (3/3)

- **Other**
  - VDE-AR-E 2842-61 - Development and trustworthiness of autonomous/cognitive systems
  - AS6983 - Process Standard for Development and Certification/Approval of Aeronautical Safety-Related Products Implementing AI
  - EUROCAE WG114 - to prepare technical standards, guides and any other material required to support the development of systems and the certification of aeronautical systems implementing AI-technologies

TR 5469

- Currently in DTR stage
- Developed in liaison with MT 61508 (AIFS task force of MT 61508-3)
- TR expected to be published in Summer 2023

ISO/IEC TR 5469:202x(E)
ISO/IEC JTC 1/SC 42/WG 3

Secretariat: ANSI

Artificial intelligence — Functional safety and AI systems

Draft Technical Report stage

Warning for WDs and CDs

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

TR 5469 contents

DTR stage, November 2022

Contents

TR 5649 scope

- The DTR 5469 standard describes the properties, related risk factors processes relating to:

01

Use of AI inside a safety related function to realize the functionality

02

Use of non-AI safety related functions to ensure safety for an AI controlled equipment

03

Use of AI systems to design and develop safety related functions

TR 5469 three stages realization principle

Data acquisition

↓

Knowledge induction

↓

Processing and generation of outputs

Desirable Properties → **Topics** → **Methods & Techniques** → **Acceptance Criteria / KPI**

Desirable Properties

e.g.

- Specifiability
- Domain shift
- Verifiability
- Robustness
- Interpretability
- Explainability

Topics

e.g. for verifiability

- How is the neural network performance assessed?
- How to determine when verification is complete?

Methods & Techniques

e.g. for performance

- Definition of what constitutes a “correct” output by the network.
- Definition of what range of inputs is evaluated

Acceptance Criteria / KPI

e.g. for performance

- Pixel level KPIs
- Image level KPIs
- Sequence level KPIs
- Dataset level KPIs

AI lifecycle vs functional safety lifecycle

- ISO/IEC 22989 describes a high-level lifecycle model of AI systems
- ISO/IEC 5338 defines the lifecycle processes of AI systems
- IEC 61508 describes the functional safety lifecycle

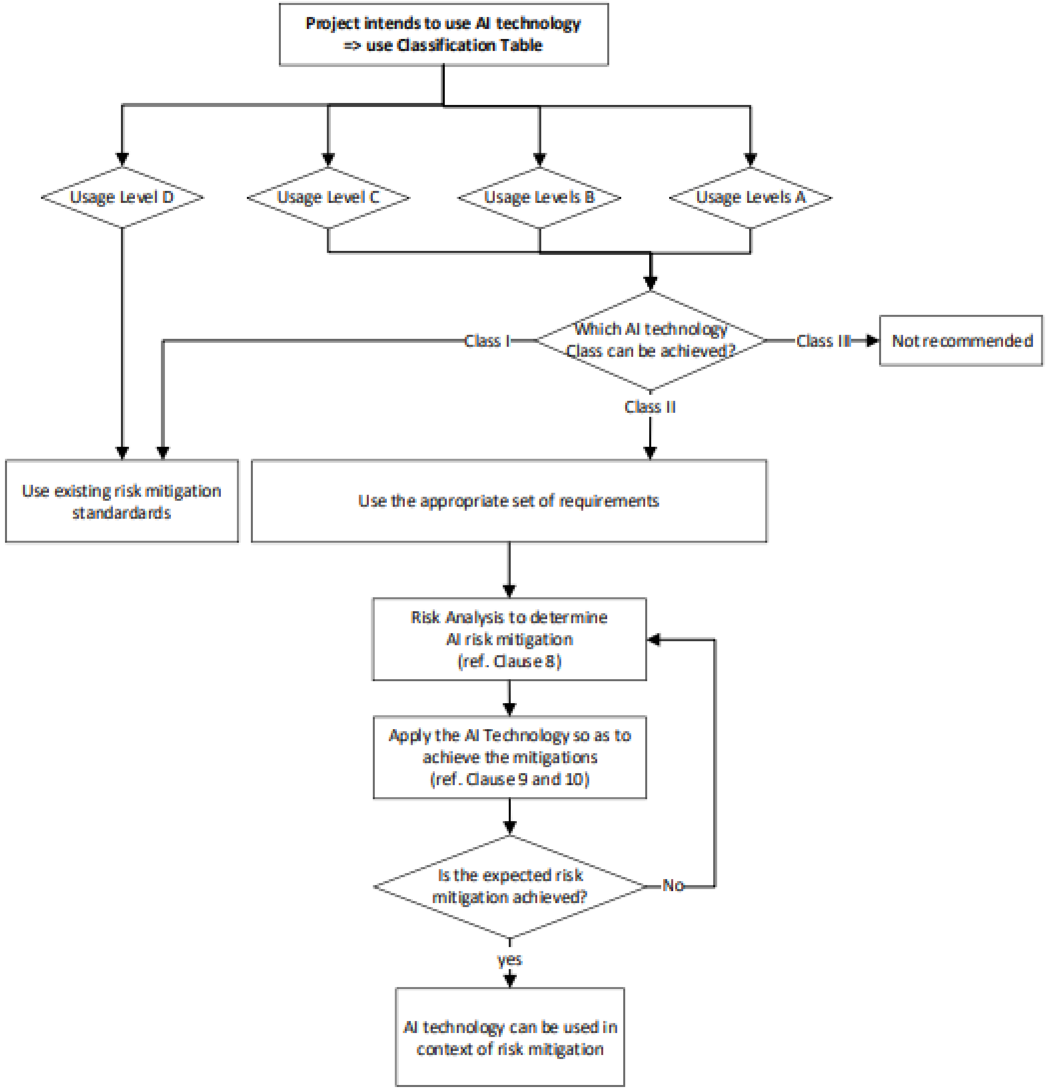
An example of mapping between ISO/IEC 5338 and the IEC 61508 series is provided in Annex D.

TR 5469 hierarchy of technology elements

- Application graph, machine learning framework
- Machine learning graph compiler, machine learning model
- Libraries of calculation operands, set of calculations
- Executable machine code, compiler
- Computational HW

TR 5469 AI classification scheme

DTR stage, November 2022



DTR stage, November 2022

AI Technology Class => AI application and usage level	AI technology Class I	AI technology Class II	AI technology Class III
Usage Level A1 (1)	Application of risk reduction concepts of existing functional safety International Standards possible	Appropriate set of requirements (5)	Not recommended
Usage Level A2 (1)		Appropriate set of requirements (5)	
Usage Level B1 (1)		Appropriate set of requirements (5)	
Usage Level B2 (1)		Appropriate set of requirements (5)	
Usage Level C (1,3)		Appropriate set of requirements (5)	
Usage Level D (2)	No specific functional safety requirements for AI technology, but application of risk reduction concepts of existing functional safety International Standards (4)		

1 Static (offline) (during development) teaching or learning only

2 Dynamic (online) teaching or learning possible

3 AI techniques clearly providing additional risk reduction and whose failure is not critical to the level of acceptable risk.

4 Additionally, other safety aspects (not being addressed with functional safety methods) can possibly be adversely affected by AI usage.

5 The appropriate set of requirements for each usage level can be established in consideration of Clauses 8, 9, 10 and 11. Examples are provided in Annex B.

TR 5469 architectural patterns for safe AI

Redundancy	Supervision	Back-up
• ML redundancy with output voters or aggregators	• A safe subset of the action space can be determined using a supervisor function with constraints or limits	• Safe (suboptimal) back-up function to the ML component can be designed with ”non-AI” techniques. • The back-up action allows the use of detection methods to switch the output when unsafe conditions are detected.

TR 5469 and diversity of AI technologies

Redundancy is combined with diversity to reduce the likelihood of systematic failures during development.

This is related to multiple AI technologies exhibiting the same behaviour, but implemented:

- by different teams;
- using separate labelling rules;
- using different problem formulations;
- using different training data;
- executing on diverse hardware (also valid for non-AI technology specific failure modes);
- with diversity of sensing;
- with diversity of self-check or self-validation methods;
- with diversity of AI technology itself.

Virtual and physical testing

- An effective and objective way to demonstrate a system's performance is via **virtual testing or simulation**, where a curated set of well-chosen stress-test scenarios can be exercised during the qualification and certification activities.
- **Physical tests** also have their place as a tangible way to correlate simulation results, validate KPIs and uncover unknown unknowns.
  - Physical tests are far more limited than simulation in their ability to probe the domain space due to cost and time limitations but do test some aspects that are difficult to emulate in a simulation, for example, the effect of hardware delays on feedback loops and cascade effects.

Thank you !