Mehr Wert.
Mehr Vertrauen.

Artificial Intelligence and Functional Safety

A summary of the current challenges

Volker Schneider
TÜV SÜD Rail GmbH
Rail Automation

27. May 2021

Our Expert

Volker Schneider
Expert Functional Safety at TÜV SÜD

Background:

- Volker worked for 5 years as research assistant at TU München.
- He did his Ph.D. in the field of trajectory generation for integrated flight guidance.
- In this function he was developing model based software for a manned experimental aircraft under consideration of safety related aspects like traceability, testability and architectural aspects related to model based development.
- Volker joined TÜV SÜD in 2017 as functional safety assessor for several domains regarding the standards IEC 61508, and ISO 26262.

Where is AI discussed International….

ISO/IEC JTC 1/SC 42 [1]

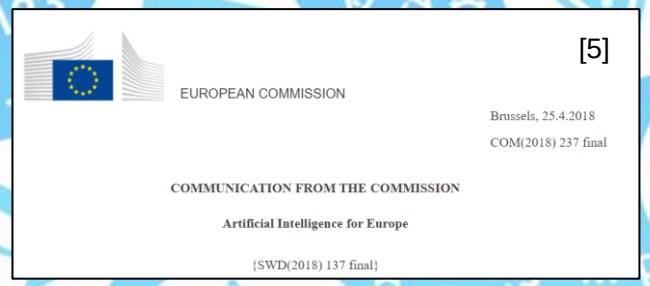
WG1: Standardization of Terms
e.g. ISO/IEC 22989: AI – Concept and Terminology
WG2: Data + Data quality
e.g. ISO/IEC 24688 Information Technology – Artificial Intelligence – Process management framework for Big data analytics
WG3: Trustworthiness
ISO/IEC AWI TR 5469 Artificial intelligence — Functional safety and AI systems
WG4: Use Cases
e.g. ISO/IEC 24030 Information Technology – Artificial Intelligence – Use Cases
WG5: Calculation Aspects+ Characteristics
e.G ISO/IEC 24372 Information Technology – Artificial Intelligence – Overview of computational approaches for AI Systems

IEEE P7000 [2]

IEEE P7000™ - Standard for Model Process for Addressing Ethical Concerns During System Design
IEEE P7001™ - Standards for Transparency of Autonomous Systems
IEEE P7002™ - Standard for Data Privacy Process
…
IEEE P7009™ - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems
…
IEEE P7014™ - Standard for Ethical considerations in Emulated Empathy in Autonomous and Intelligent Systems

AI on european level

- **European Comisssion**
  - Several publications regarding AI topic
  - Latest from April 2021
- **European Comitee for Standardization CEN**:
  - Supports the activities of ISO/IEC JTC1 SC42
  - Established a Focus Group on Artificial Intelligence


EUROPEAN COMMISSION

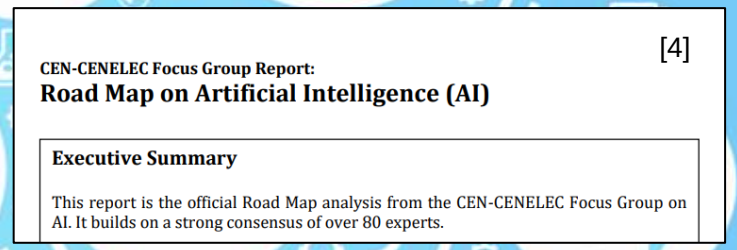
[5]

Brussels, 25.4.2018
COM(2018) 237 final

COMMUNICATION FROM THE COMMISSION

Artificial Intelligence for Europe

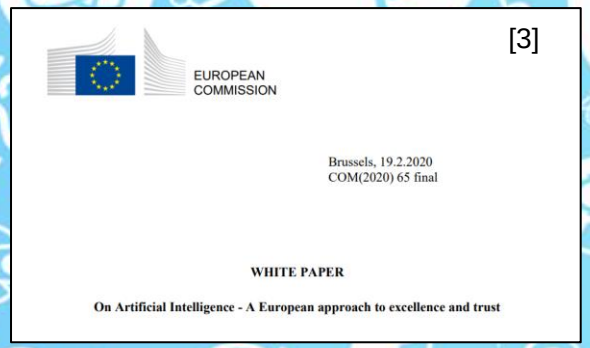
{SWD(2018) 137 final}


[4]

CEN-CENELEC Focus Group Report:
Road Map on Artificial Intelligence (AI)

Executive Summary

This report is the official Road Map analysis from the CEN-CENELEC Focus Group on AI. It builds on a strong consensus of over 80 experts.

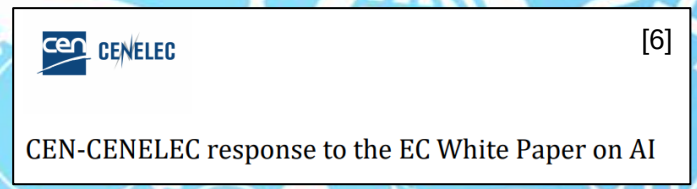

EUROPEAN COMMISSION

[3]

Brussels, 19.2.2020
COM(2020) 65 final

WHITE PAPER

On Artificial Intelligence - A European approach to excellence and trust


CENELEC

[6]

CEN-CENELEC response to the EC White Paper on AI


EUROPEAN COMMISSION

[9]

Brussels, 21.4.2021
COM(2021) 206 final

2021/0106 (COD)

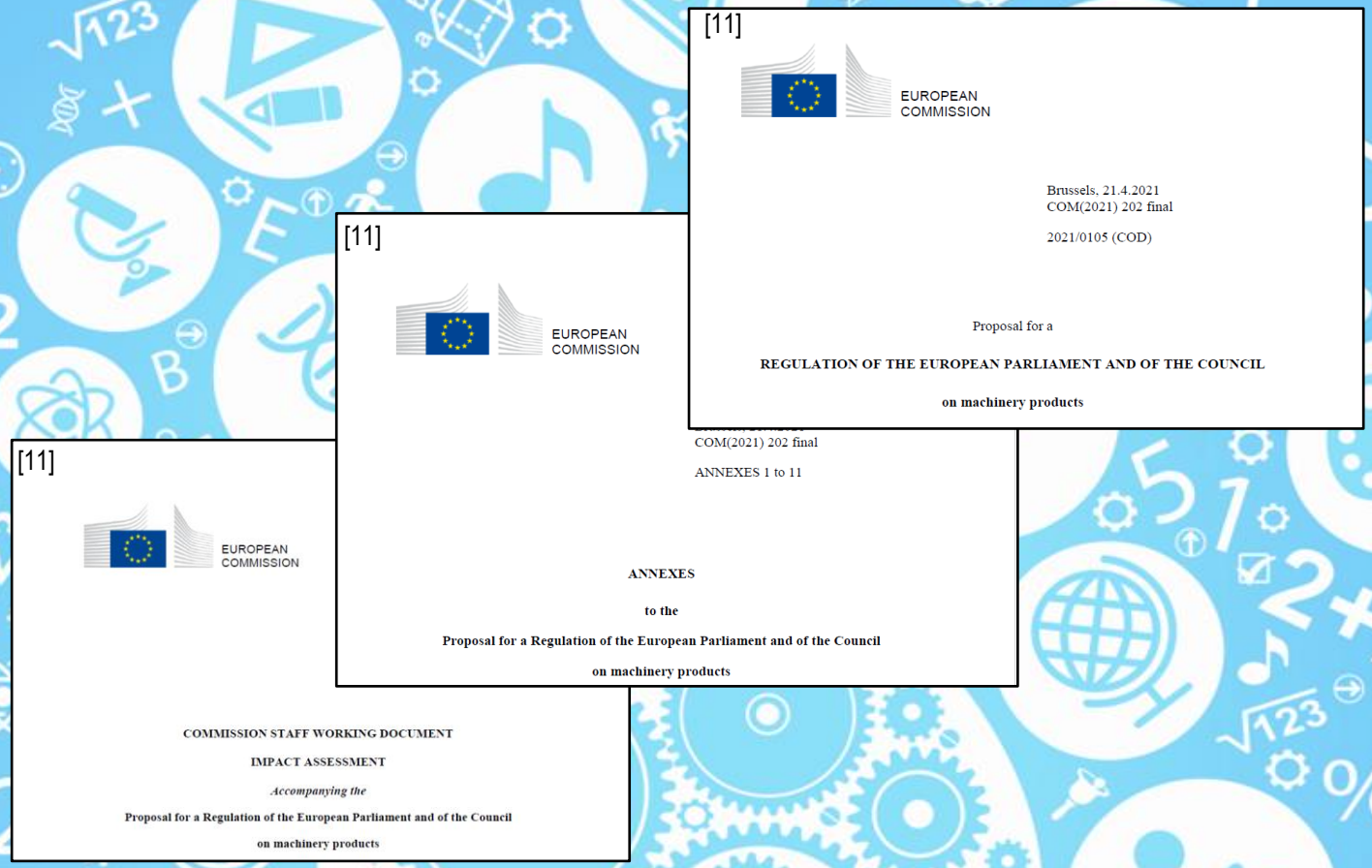
Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

Machinery directive

- Proposal recently published
- Annex document: Relation also to AI / autonomous levels
- Sections to consider:
  - 1. c)
  - 1.1.6 e) and f)
  - 1.2.1
  - 1.3.7.

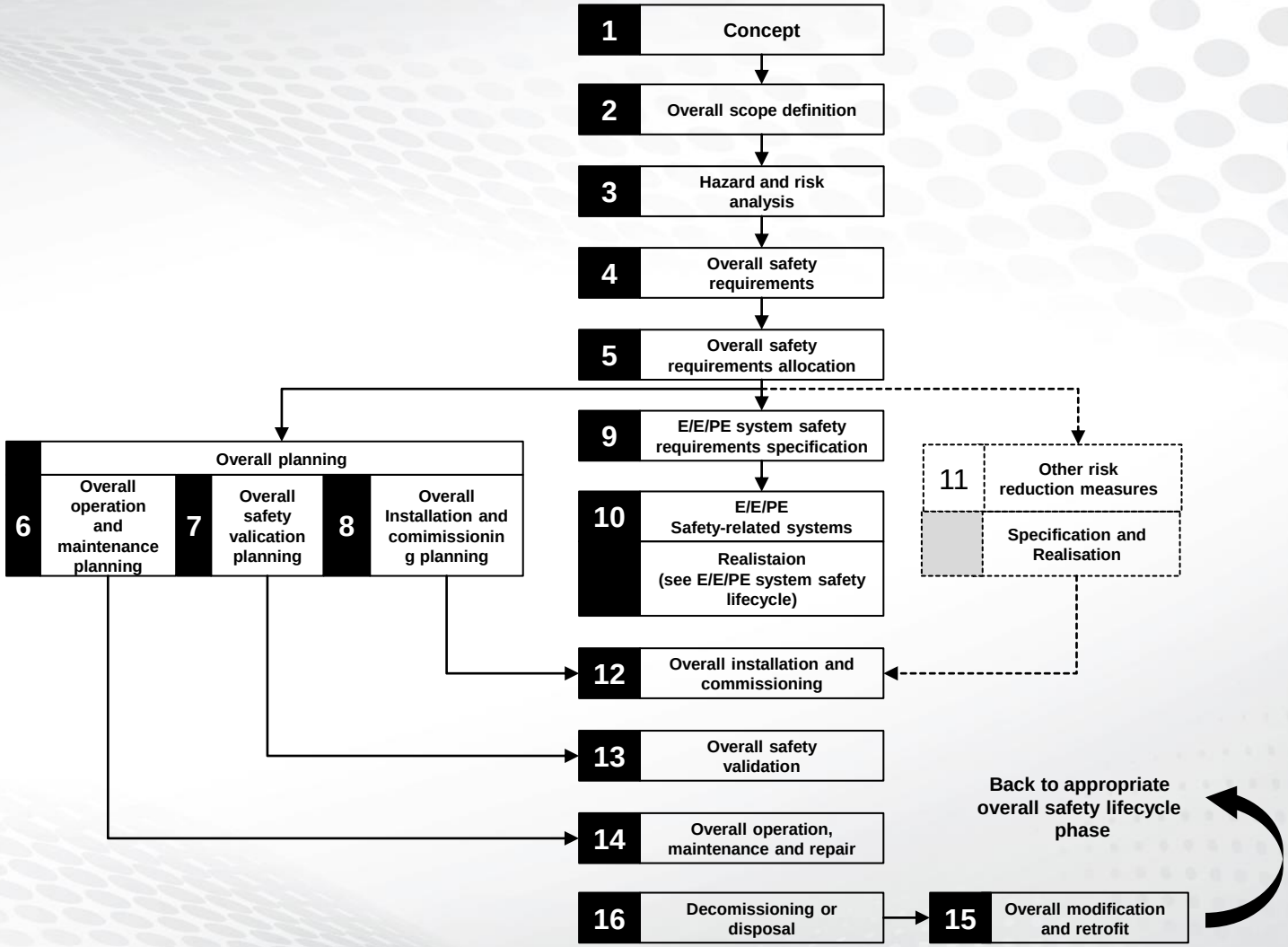


Example: Annex 1.c)
In this respect, where the machinery product integrates an artificial intelligence system, the machinery risk assessment **shall consider the risk assessment for that artificial intelligence system that has been carried out pursuant to the Regulation …** of the European Parliament and of the Council+ on a European approach for Artificial Intelligence+[1]; .

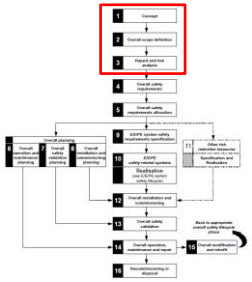
Development lifecycle

- Content: Specific aspects of the development lifecycle according to IEC 61508:2010

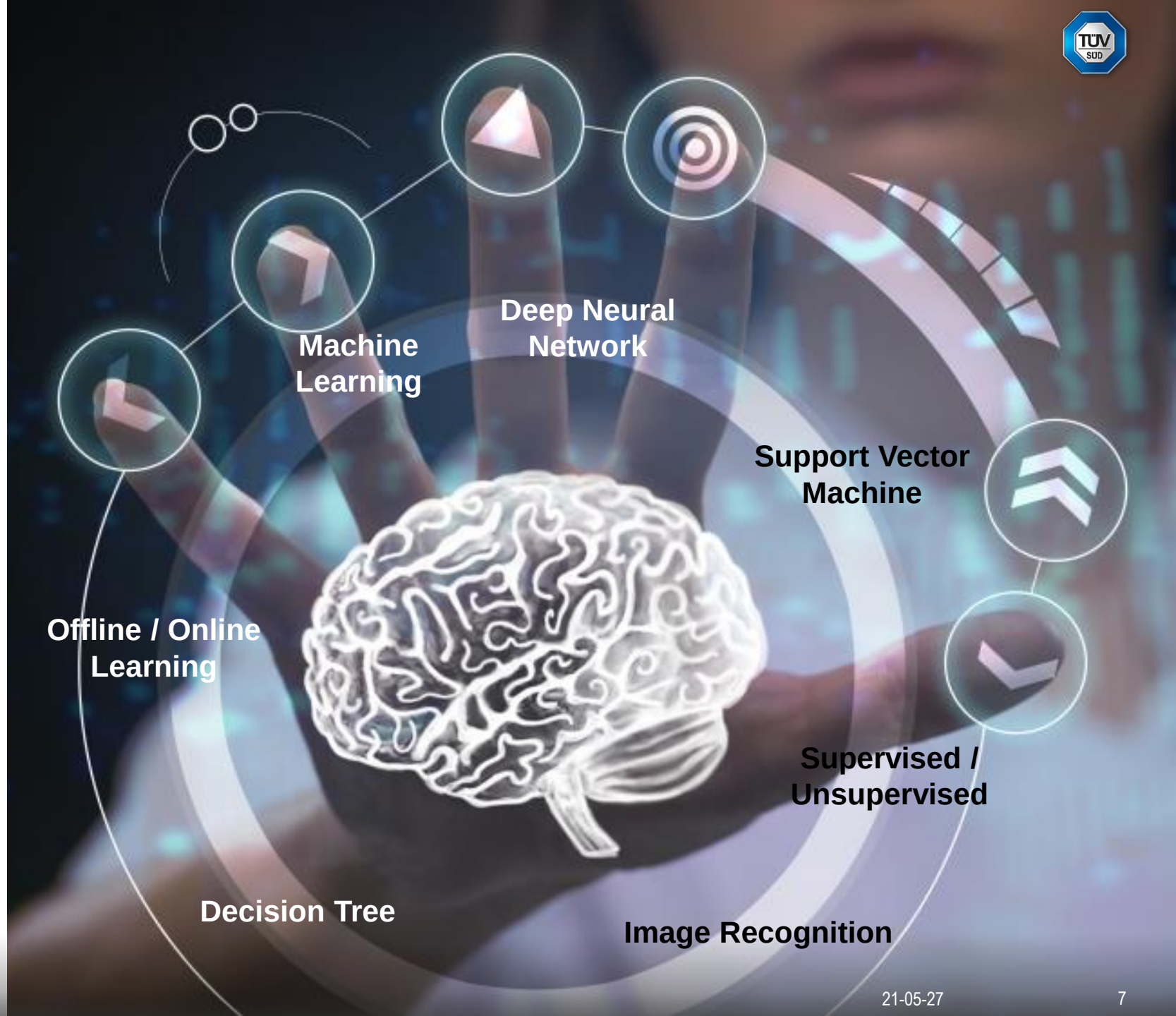
- Evaluation regarding specific AI related topics



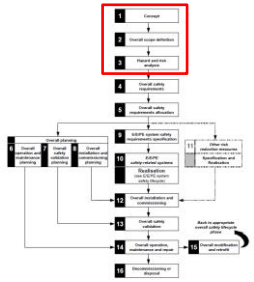
Concept / Risk Analysis



- ISO/IEC 2382:2015: "***Artificial Intelligence (AI):*** Capability of a function unit to perform functions that are generally associated with human intelligence such as reasoning and learning"
- Term „AI" often misinterpreted.
- It is important to clearly point out what exactly is meant.



Concept / Risk Analysis

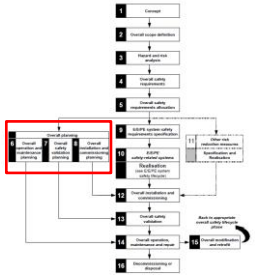


- Known risk analysis approaches on item level also applicable.
- **BUT:** Are extisting standards adequate for AI handling?
- TR 5469: Discussion regarding AI algorihm classification

	AI Class I Evaluation possible	**AI Class II** Evaluation partly not possible, but applicable with addtitional measures	**AI Class III** Evaluation not sufficiently possible, additional measures not sufficiently applicable
Usage Level A AI in E/E/PE-system for diagnosis (A2) or for control (A1)	Existing standards for risk mitigation measures regarding Functional Safety are applicable	***Area of new methods and measures to specify***	Use not recommended
Usage Level B1 or B2 AI for the development of an E/E/PE-system as a support tool (B2) or as a validation tool (B1)			
Usage Level C AI not safety relevant but with interference to safety system			
Usage Level D AI not safety relevant and free from interference to safety system	Existing risk mitigation measures regarding Functional Safety are applicable		

Based on [10]

Functional Safety Management

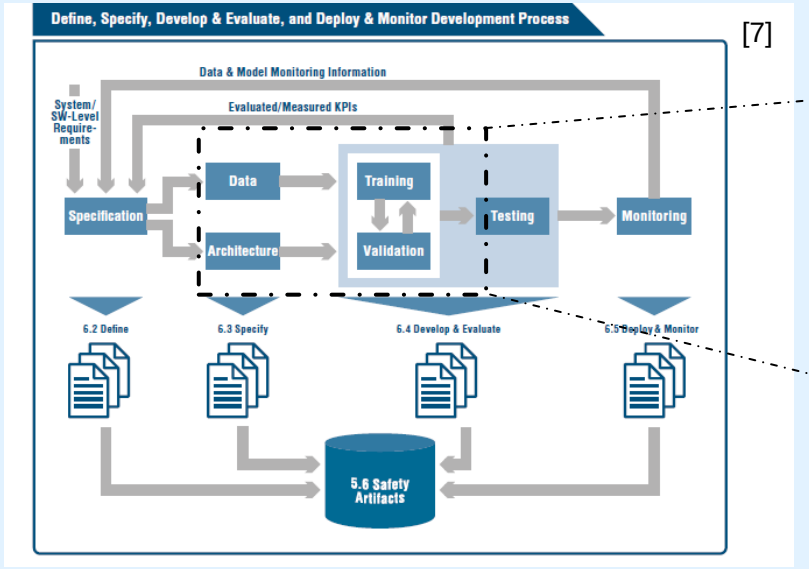


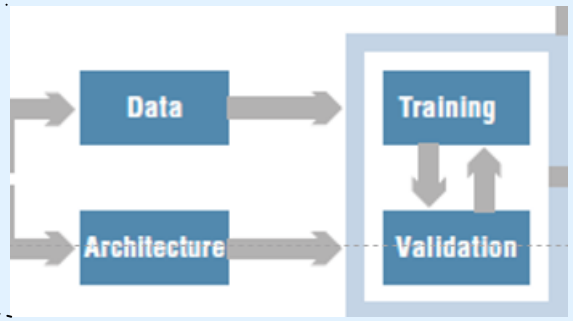
- Classical Safety Management elements (e.g. configuration management , change management, safety planning,…) are applicable

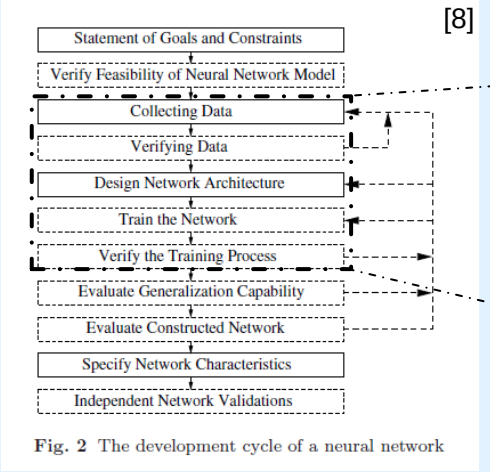
- **BUT:** for AI development differ from conventional development

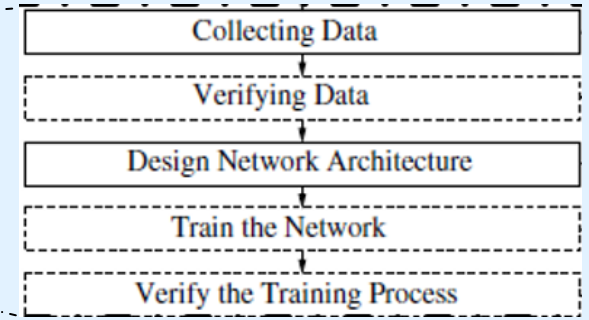
  - Data handling
  - Model training

See also:
https://wiki.eclipse.org/images/0/0e/WhitePaper_Process_considerations-A_reliable_AI_data_labeling_process.pdf

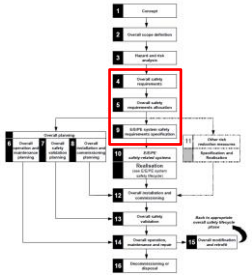






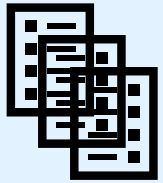


Requirements Specification

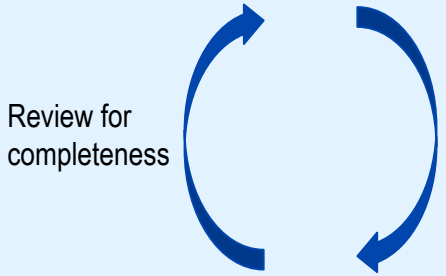


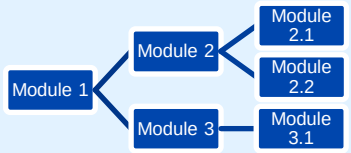
- Classic approach: Behavior fully specified by requirements.
- AI /ML: Behavior specified by dedicated test points.
- Training Data quality essential
- **Question: Is the specification „complete enough“?**

Classical Approach



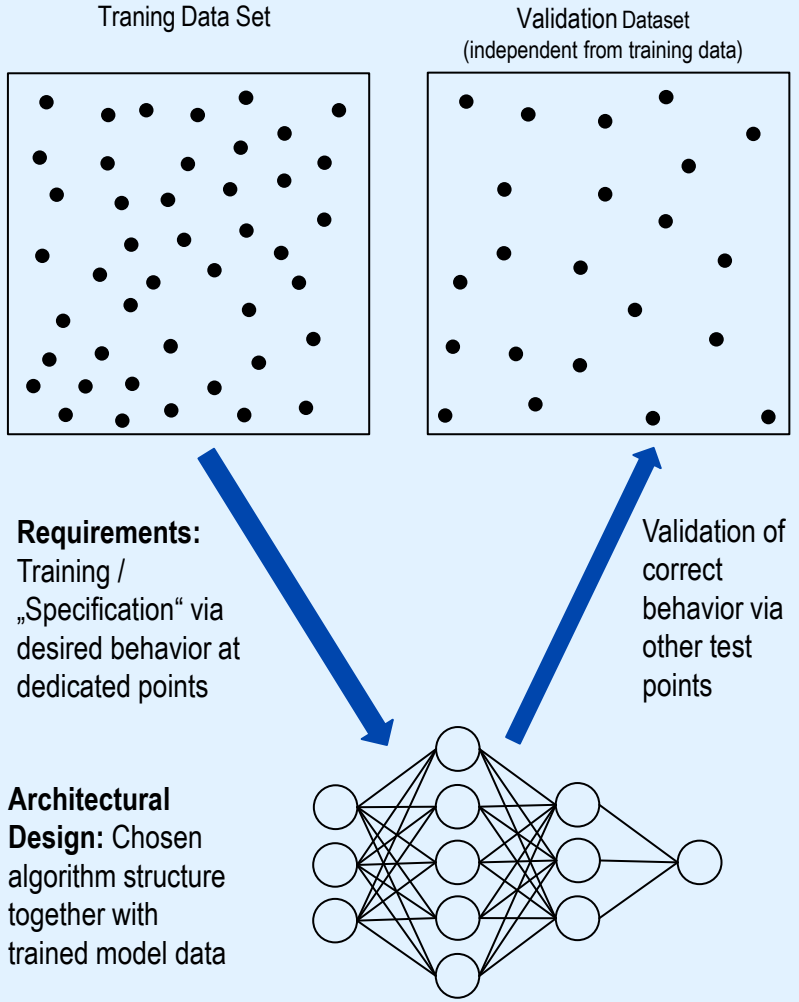
Requirements:
WHAT shall be done?



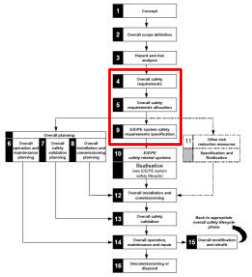


Architectural Design:
Description of HOW the system fulfils the requirements

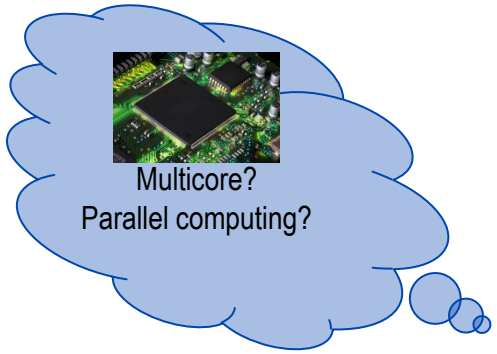
Machine Learning Approach

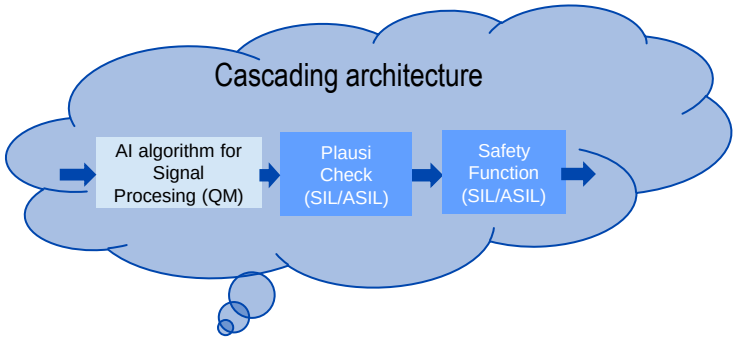


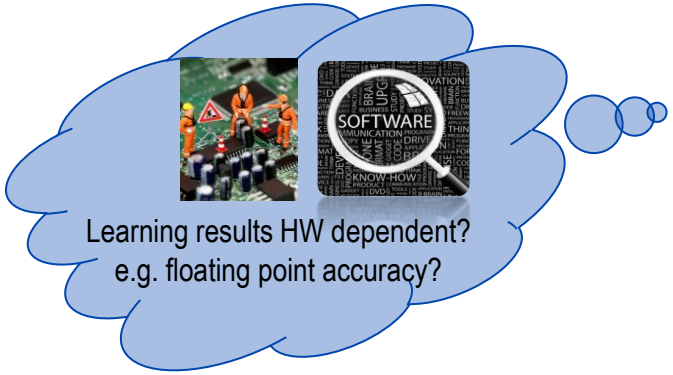
Design

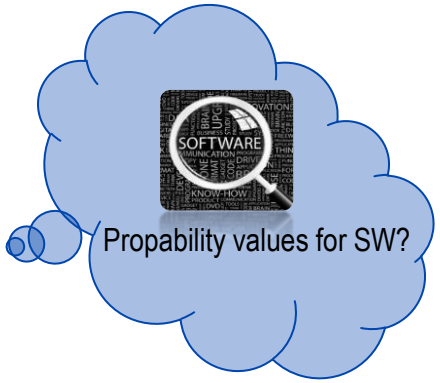


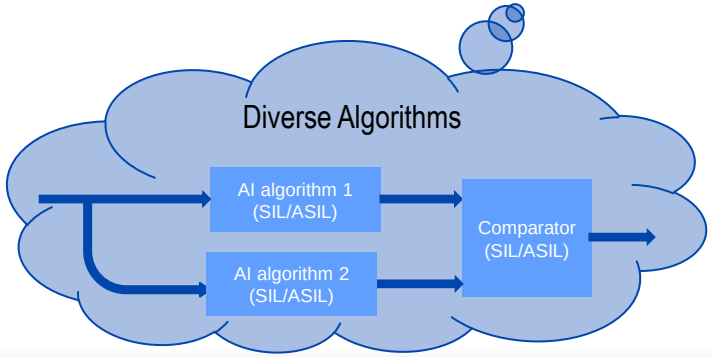
- Questions and topics to discuss:
  - Deep neural network as a „deterministic“ algorithm?
  - HW handling: AI topics do consider?
  - Which Architecture to be used?
    i. Supervisor
    ii. Diverse Algorithms
    iii. Cascading architecture

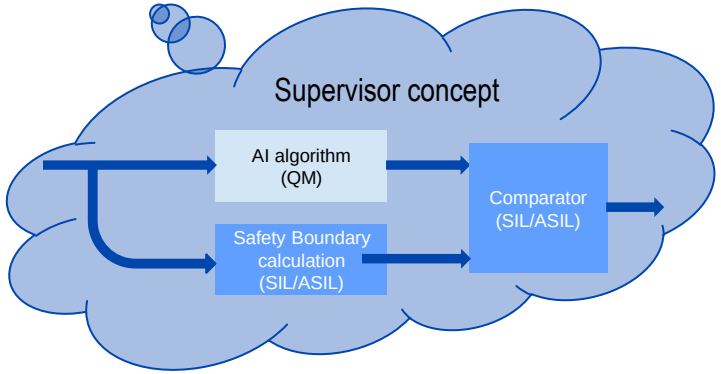




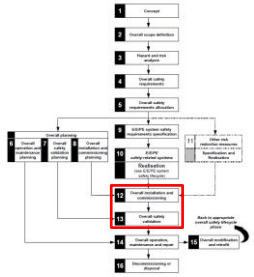






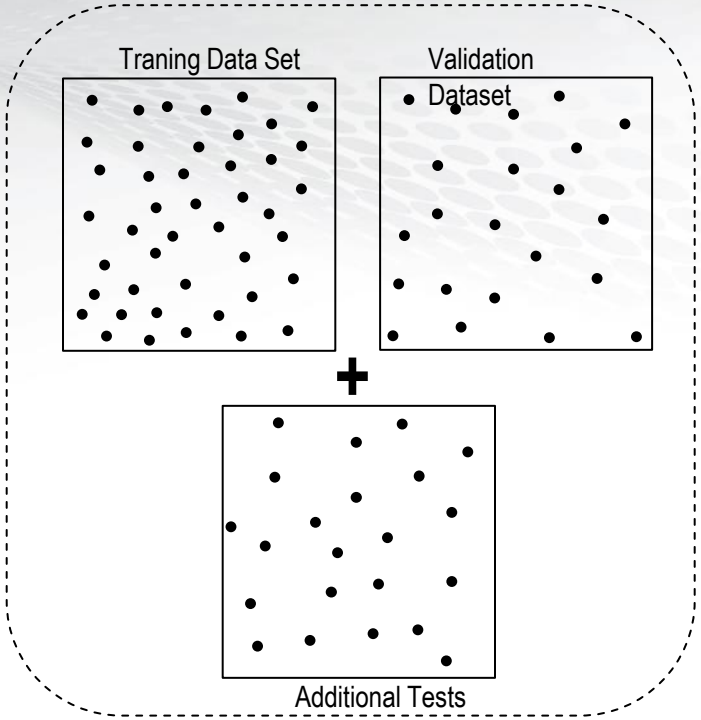


Requirements Verification



- Challenge: How to test a system without a „complete“ specification

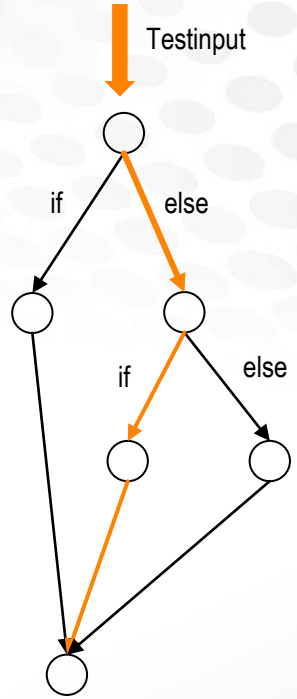
- Does the structural coverage provide the similar statement compared to classic SW development?



=

Test Space

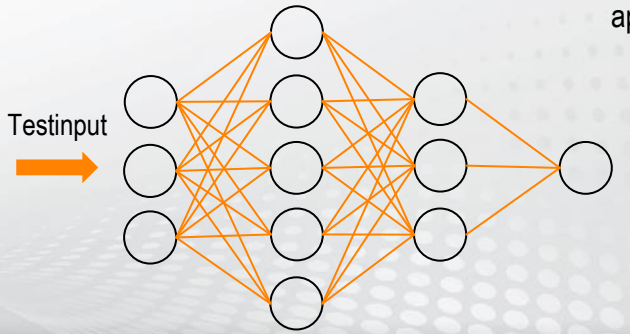
Can we make reliable statements regarding the behavior in the remaining white space (Equivalence classes)?



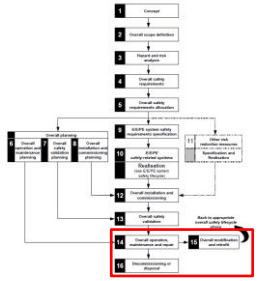
Classical structural coverage approach

≠

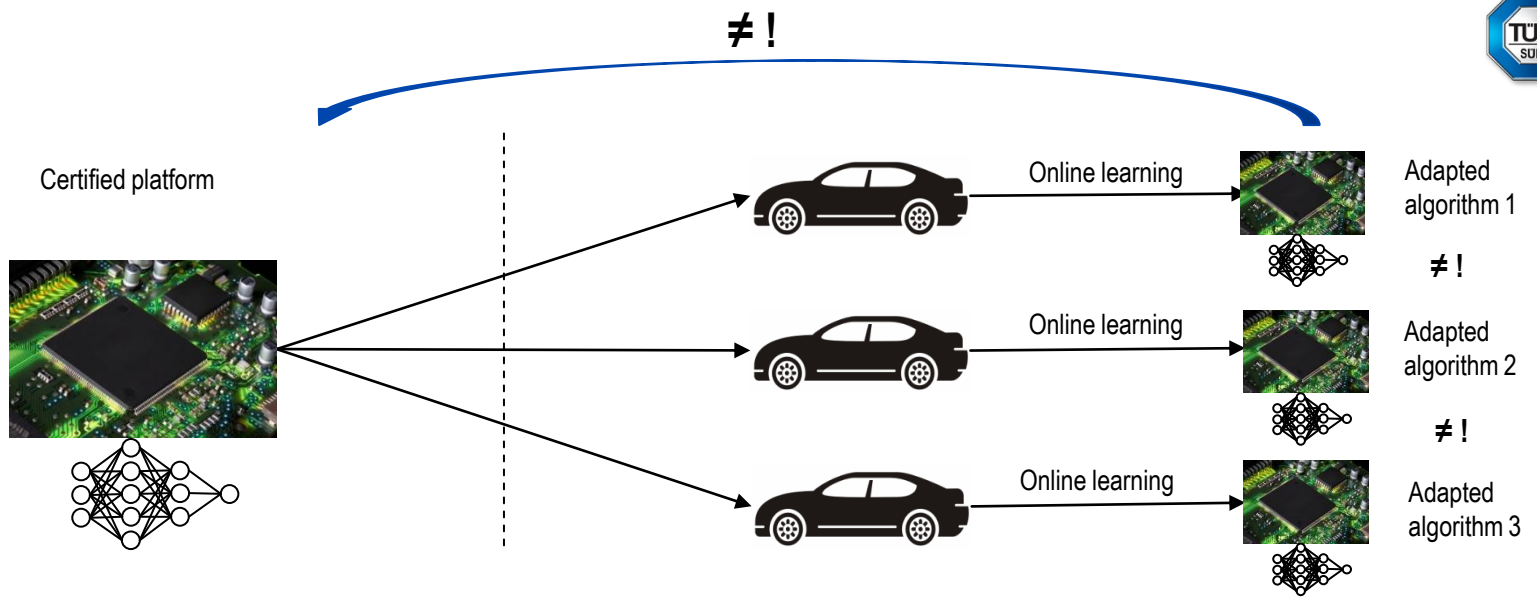
AI Testing approach



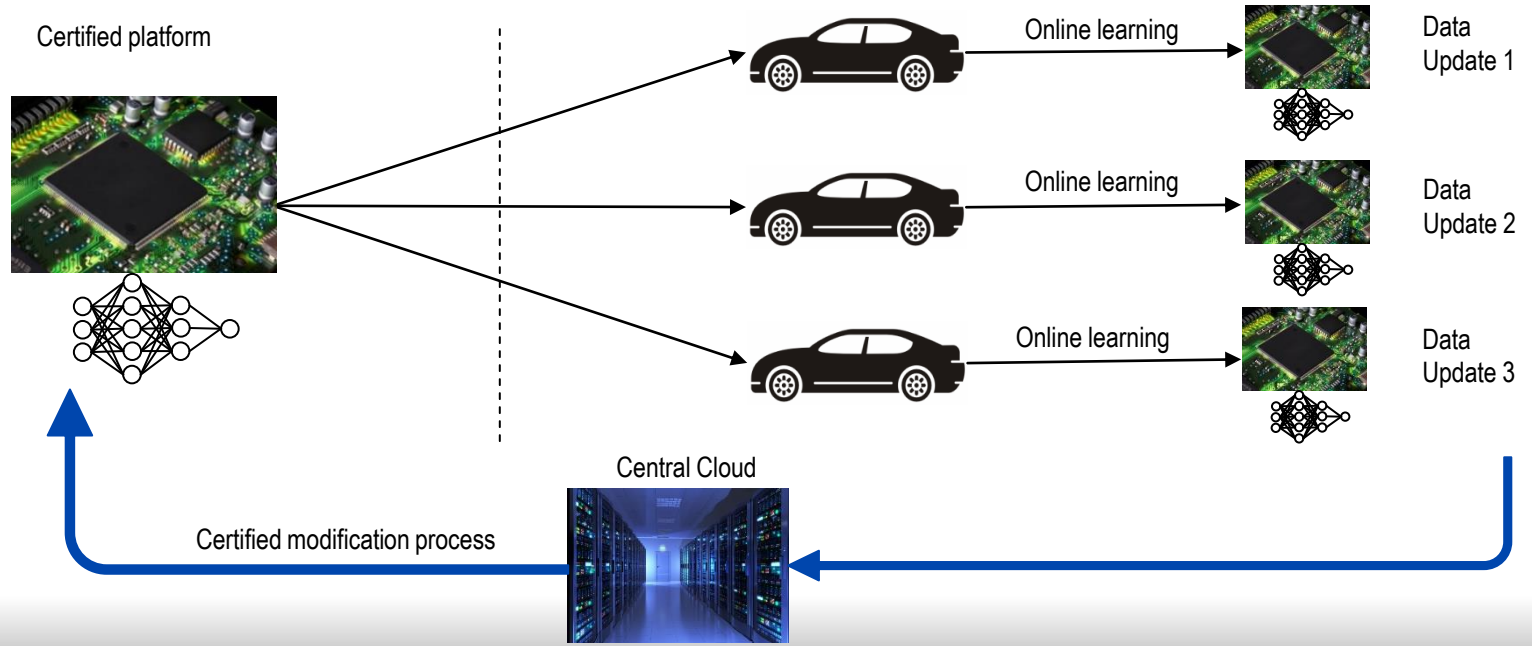
Field behavior



- Is the system learning during field operation?
- What is then the result of an assessment?
- Is it possible to establish online „quality measures“ for an AI system?
- For specific approaches, also a central „online learning“ could be a solution (e.g. central map management)



Discussed approach for specific use cases:



Thank you for your attention!

References

[1] https://www.iso.org/committee/6794475.html
[2] https://standards.ieee.org/initiatives/artificial-intelligence-systems/standards.html#p7000
[3] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
[4] https://ftp.cencenelec.eu/EN/EuropeanStandardization/Sectors/AI/CEN-CLC_FGR_RoadMapAI.pdf
[5] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN
[6] https://ftp.cencenelec.eu/EN/News/PolicyOpinions/2020/CEN-CLC_AI_FG_White-Paper-Response_Final-Version_June-2020.pdf
[7] see e.g. https://www.daimler.com/innovation/case/autonomous/safety-first-for-automated-driving-2.html
[8] J. Schumann, Y. Liu (Eds.): Appl. of Neural Networks in High Assur. Sys., SCI 268, pp. 1–19
[9] https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence
[10] H. Laible: „ Eine KI Klassifikation für Safety”, Safe.Tech 2021
[11] https://ec.europa.eu/docsroom/documents/45508?locale=fi