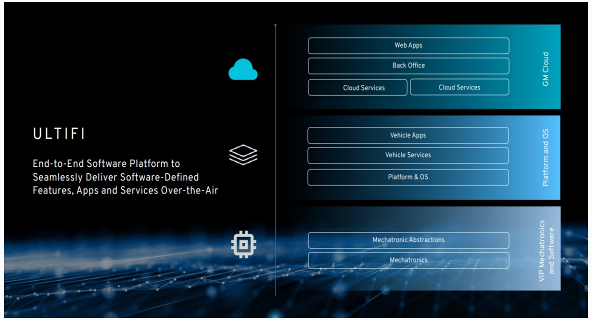
Overview of Emerging Safety Standards:
ADS & ARTIFICIAL INTELLIGENCE

Ramesh S

General Motors R&D

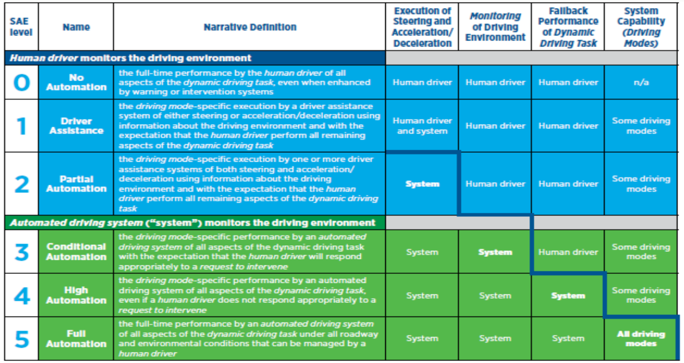
NEXT GEN VEHICLES

- SW Defined Vehicles
  - Functionality defined in the hands of the customer
  - Continuous Improvement/Continuous Deployment
  - Incremental Functionality transported Over-The-Air (OTA)
- Automated Driving System
  - SAE Level 3 and 4
  - Complex features
  - Perception & Planning
- AI and Machine Learning Enabled
  - Deep Neural Networks for perception tasks
  - Reinforcement Learning for planning

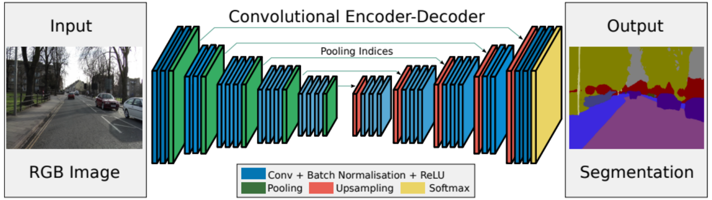


ULTIFI

End-to-End Software Platform to
Seamlessly Deliver Software-Defined
Features, Apps and Services Over-the-Air

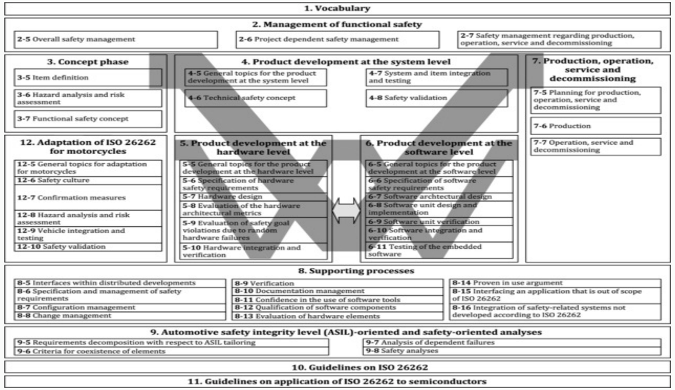


SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring* of Driving Environment	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
***Human driver* monitors the driving environment**						
0	**No Automation**	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	**Driver Assistance**	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	**Partial Automation**	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes
***Automated driving system* ("system") monitors the driving environment**						
3	**Conditional Automation**	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes
4	**High Automation**	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes
5	**Full Automation**	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**



Safety Engineering and Road Vehicles

- Has a long history
- Two Standards and subsequent revisions
  - ISO 26262: Functional Safety
  - ISO 21448: Safety of the Intended Function (SoTIF)
- Functional Safety
  - Safety under random failures of HW and systematic failures of SW
  - ASIL and elaborate Design, Verification & Validation Processes
- SoTIF
  - Safety in spite of functional insufficiency or misuse
  - Trigger conditions and Acceptance Criteria
  - Scenario based testing
- Both standards attempt to accommodate the new way of building automotive software – CI/CD



1. Vocabulary

2. Management of functional safety

2-5 Overall safety management

2-6 Project dependent safety management

2-7 Safety management regarding production, operation, service and decommissioning

3. Concept phase

3-5 Item definition

3-6 Hazard analysis and risk assessment

3-7 Functional safety concept

4. Product development at the system level

4-5 General topics for the product development at the system level

4-6 Technical safety concept

4-7 System and item integration and testing

4-8 Safety validation

7. Production, operation, service and decommissioning

7-5 Planning for production, operation, service and decommissioning

7-6 Production

7-7 Operation, service and decommissioning

12. Adaptation of ISO 26262 for motorcycles

12-5 General topics for adaptation for motorcycles

12-6 Safety culture

12-7 Confirmation measures

12-8 Hazard analysis and risk assessment

12-9 Vehicle integration and testing

12-10 Safety validation

5. Product development at the hardware level

5-5 General topics for the product development at the hardware level

5-6 Specification of hardware safety requirements

5-7 Hardware design

5-8 Evaluation of the hardware architectural metrics

5-9 Evaluation of safety goal violations due to random hardware failures

5-10 Hardware integration and verification

6. Product development at the software level

6-5 General topics for the product development at the software level

6-6 Specification of software safety requirements

6-7 Software archtectural design

6-8 Software unit design and implementation

6-9 Software unit verification

6-10 Software integration and verification

6-11 Testing of the embedded software

8. Supporting processes

8-5 Interfaces within distributed developments

8-6 Specification and management of safety requirements

8-7 Configuration management

8-8 Change management

8-9 Verification

8-10 Documentation management

8-11 Confidence in the use of software tools

8-12 Qualification of software components

8-13 Evaluation of hardware elements

8-14 Proven in use argument

8-15 Interfacing an application that is out of scope of ISO 26262

8-16 Integration of safety-related systems not developed according to ISO 26262

9. Automotive safety integrity level (ASIL)-oriented and safety-oriented analyses

9-5 Requirements decomposition with respect to ASIL tailoring

9-6 Criteria for coexistence of elements

9-7 Analysis of dependent failures

9-8 Safety analyses

10. Guidelines on ISO 26262

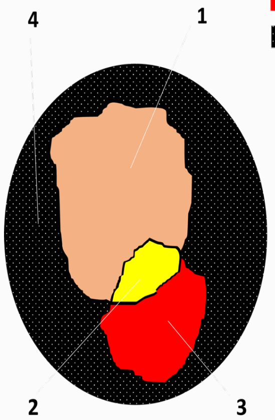
11. Guidelines on application of ISO 26262 to semiconductors

🟨 Known hazardous scenarios (Area 2)

🟧 Known safe scenarios (Area 1)

🟥 Unknown hazardous scenarios (Area 3)

⬛ Unknown safe scenarios (Area 4)



	Hazardous	Safe
Known	2	1
Unknown	3	

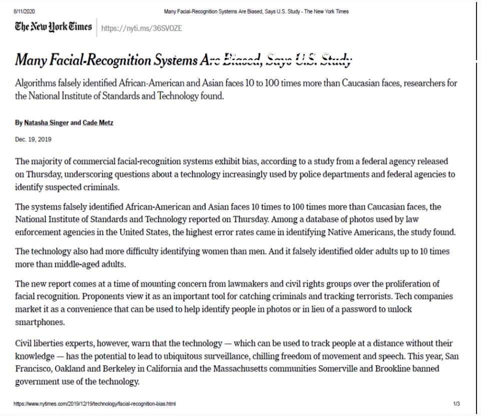
AI/ML

- AI Technology and Machine Leaning (ML) increasingly being used in vehicle applications some of which are safety-related
  - Supercruise uses MobilEye Camera which uses ML function
  - Ultracruise plans to use in-house AI/ML components
- Nature and Development and Validation of such system quite different
  - Black box, probablistic outputs
  - Data based, Training, Validation and Testing cycles
- Lack of guidelines and standards for AI system development and validation
  - OEM & Supplier specific internal guidelines
  - ISO 26262 and ISO 21448 devoted appendices to deal with AI systems
  - Several projects done and ongoing in our group

Robustness & Safety of AI systems

- AI problem is ill-posed
  - Likelihood Estimation, Stochastic
  - Incomplete (Frame Problem)
  - Bias & Uncertainty
- Predictability
  - Flounder in rare or new situations not encountered in training data set
  - Black swan issue
- Functional Safety process assumes a traditional view of development and verification
- Missing safety lifecycle assets and new assets
  - Hardly any requirements
  - Training and Test Data Sets
- Development Lifecyle for AI/ML components is non-traditional
  - Data Intensive

© BNPS.CO.UK



8/11/2020 Many Facial-Recognition Systems Are Biased, Says U.S. Study - The New York Times

The New York Times | https://nyti.ms/36SVOZE

Many Facial-Recognition Systems Are Biased, Says U.S. Study

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.

By Natasha Singer and Cade Metz

Dec. 19, 2019

The majority of commercial facial-recognition systems exhibit bias, according to a study from a federal agency released on Thursday, underscoring questions about a technology increasingly used by police departments and federal agencies to identify suspected criminals.

The systems falsely identified African-American and Asian faces 10 times to 100 times more than Caucasian faces, the National Institute of Standards and Technology reported on Thursday. Among a database of photos used by law enforcement agencies in the United States, the highest error rates came in identifying Native Americans, the study found.

The technology also had more difficulty identifying women than men. And it falsely identified older adults up to 10 times more than middle-aged adults.

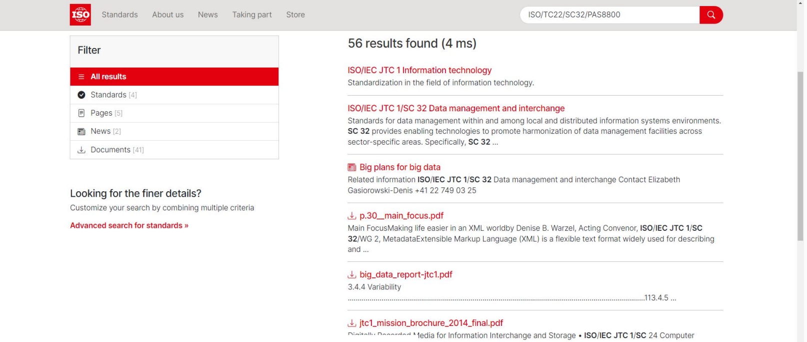
The new report comes at a time of mounting concern from lawmakers and civil rights groups over the proliferation of facial recognition. Proponents view it as an important tool for catching criminals and tracking terrorists. Tech companies market it as a convenience that can be used to help identify people in photos or in lieu of a password to unlock smartphones.

Civil liberties experts, however, warn that the technology — which can be used to track people at a distance without their knowledge — has the potential to lead to ubiquitous surveillance, chilling freedom of movement and speech. This year, San Francisco, Oakland and Berkeley in California and the Massachusetts communities Somerville and Brookline banned government use of the technology.

https://www.nytimes.com/2019/12/19/technology/facial-recognition-bias.html 1/3

The Standards Landscape for AI based systems

- Quite rich, has been an intense focus for the last few years
- More than 100 guidelines and standards in the general context have come out or under development
  - ISO/TC 22/SC32 – Electrical & Electronic components and general systems aspects
  - ISO/IEC JTC 1/SC42 Artificial Intelli
- Participating in
  - USCAR DL-SPICE Guidelines Docum
  - ISO/AWI TS 5083: Road Vehicles – for ADS – Design, V&V
  - ISP/AWI PAS 8800 Road Vehicles – and AI



ISO/TC22/SC32/PAS8800

Filter

- All results
- Standards [4]
- Pages [5]
- News [2]
- Documents [41]

Looking for the finer details?

Customize your search by combining multiple criteria

Advanced search for standards »

56 results found (4 ms)

ISO/IEC JTC 1 Information technology
Standardization in the field of information technology.

ISO/IEC JTC 1/SC 32 Data management and interchange
Standards for data management within and among local and distributed information systems environments. **SC 32** provides enabling technologies to promote harmonization of data management facilities across sector-specific areas. Specifically, **SC 32** ...

Big plans for big data
Related information **ISO/IEC JTC 1/SC 32** Data management and interchange Contact Elizabeth Gasiorowski-Denis +41 22 749 03 25

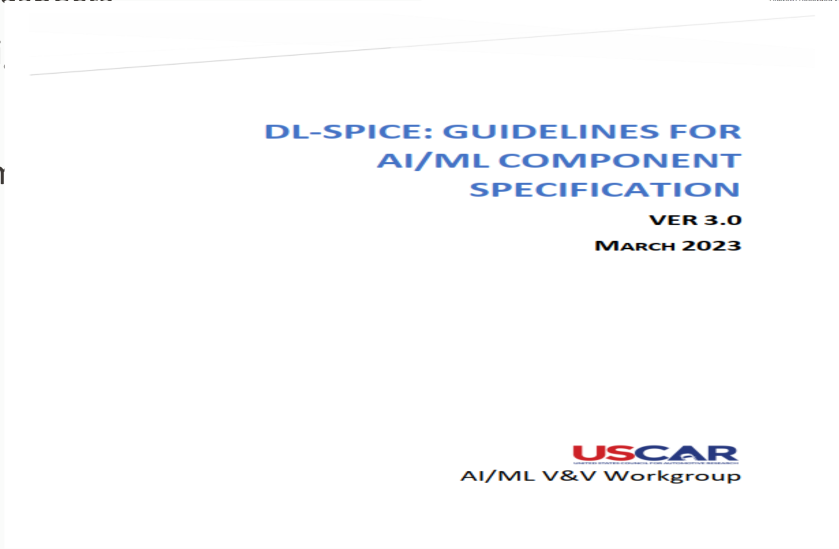
p.30__main_focus.pdf
Main FocusMaking life easier in an XML worldby Denise B. Warzel, Acting Convenor, **ISO/IEC JTC 1/SC 32**/WG 2, MetadataExtensible Markup Language (XML) is a flexible text format widely used for describing and ...

big_data_report-jtc1.pdf
3.4.4 Variability
....113.4.5 ...

jtc1_mission_brochure_2014_final.pdf
Digitally Recorded Media for Information Interchange and Storage • **ISO/IEC JTC 1/SC** 24 Computer
cessing and environmental data representation • ...

dar
lards More information about these meetings can be found on the **ISO** Meetings page.
be asked to log in. Date Month Location TC/**SC** Note 7-9 June 2022 Virtual **ISO/IEC**

2019 Health informatics — Metadata repository requirements (MetaRep)
ribes requirements for collections of metadata about data elements and their



DL-SPICE: GUIDELINES FOR AI/ML COMPONENT SPECIFICATION

VER 3.0

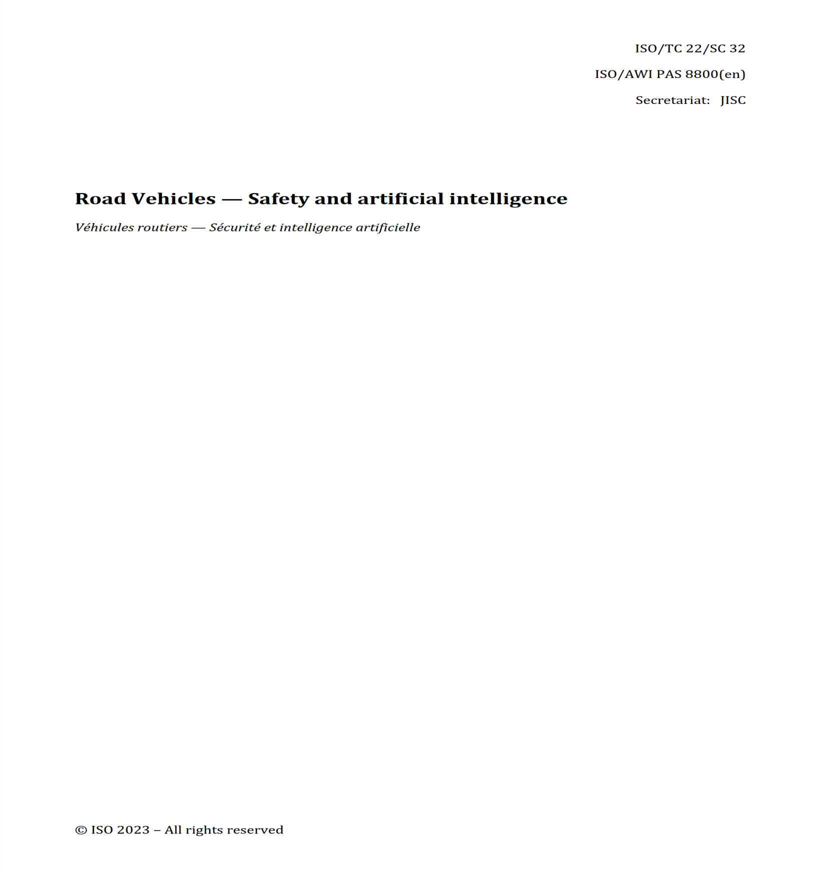
MARCH 2023

USCAR

AI/ML V&V Workgroup

ISO/PAS 8800 - Overview

- Industry-specific guidance on the use of AI/ML based systems in safety-related functions of road vehicles
  - Not restricted to specific ML techniques
  - Not restricted to ADS features
    - Annex B of ISO/TS 5083 (under development) adaptation of PAS 8800 for ADS
- Builds on guidance specified in ISO/IEC DTR 5469 (under development)
- Compatible with ISO/IS 26262 and ISO/IS 21448 (SoTIF)
- Harmonizes the concepts in Annex D.2 of ISO/IS 21448



ISO/TC 22/SC 32

ISO/AWI PAS 8800(en)

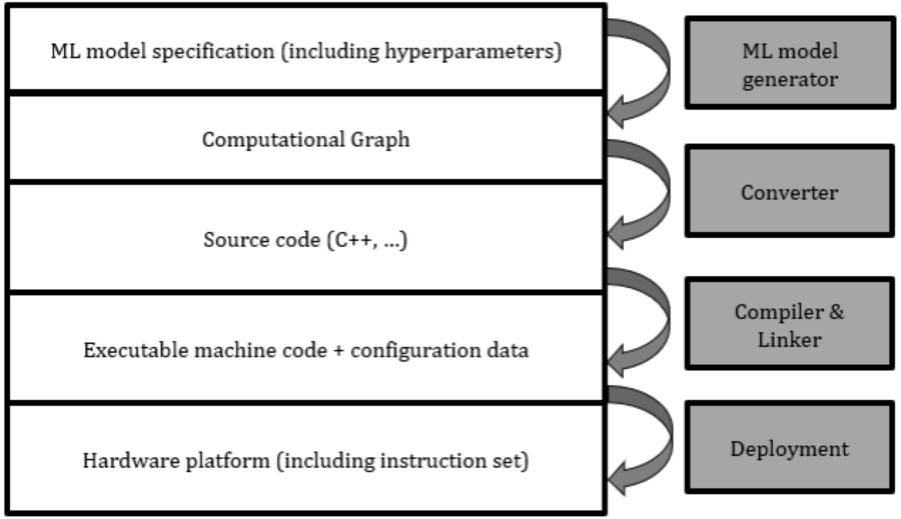
Secretariat: JISC

Road Vehicles — Safety and artificial intelligence

Véhicules routiers — Sécurité et intelligence artificielle

ISO/PAS8800 – Focus

- Items unique to the development of AI component or subsystem
  - SW and HW items along existing standards
- Safety Engineering Process along the lines of ISO 26262 and ISO 21448..
- Errors/Faults and Vehicle level Hazards
- Random HW faults and Systematic SW Faults



ML model specification (including hyperparameters)

Computational Graph

Source code (C++, ...)

Executable machine code + configuration data

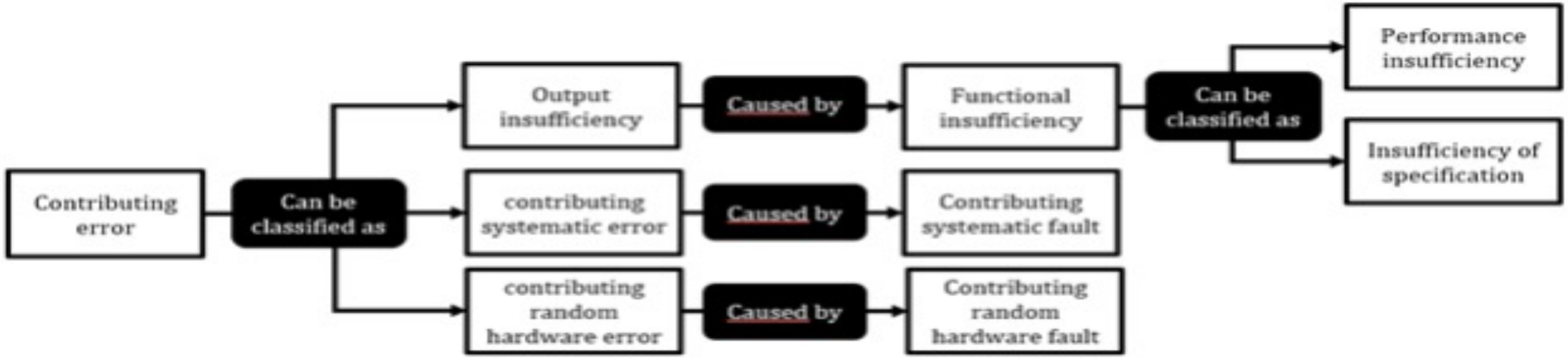
Hardware platform (including instruction set)

ML model generator

Converter

Compiler & Linker

Deployment



Contributing error

Can be classified as

Output insufficiency

contributing systematic error

contributing random hardware error

Caused by

Caused by

Caused by

Functional insufficiency

Contributing systematic fault

Contributing random hardware fault

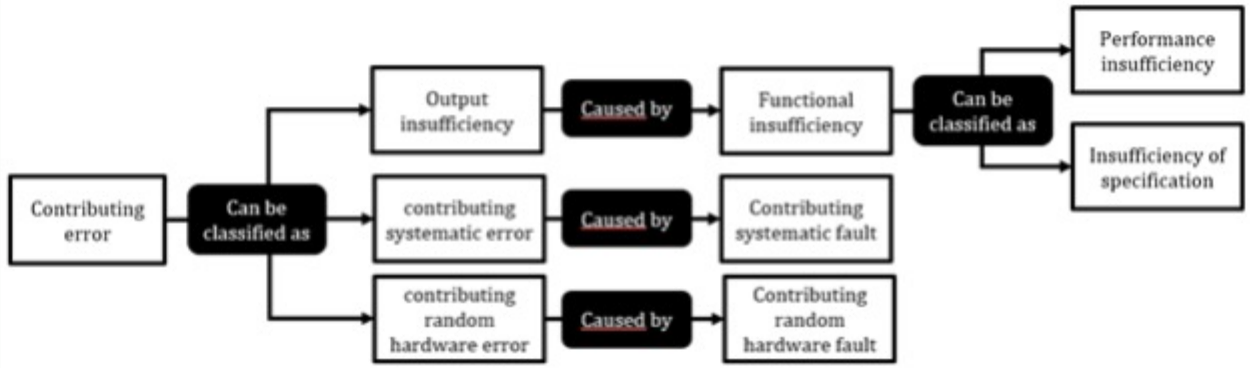
Can be classified as

Performance insufficiency

Insufficiency of specification

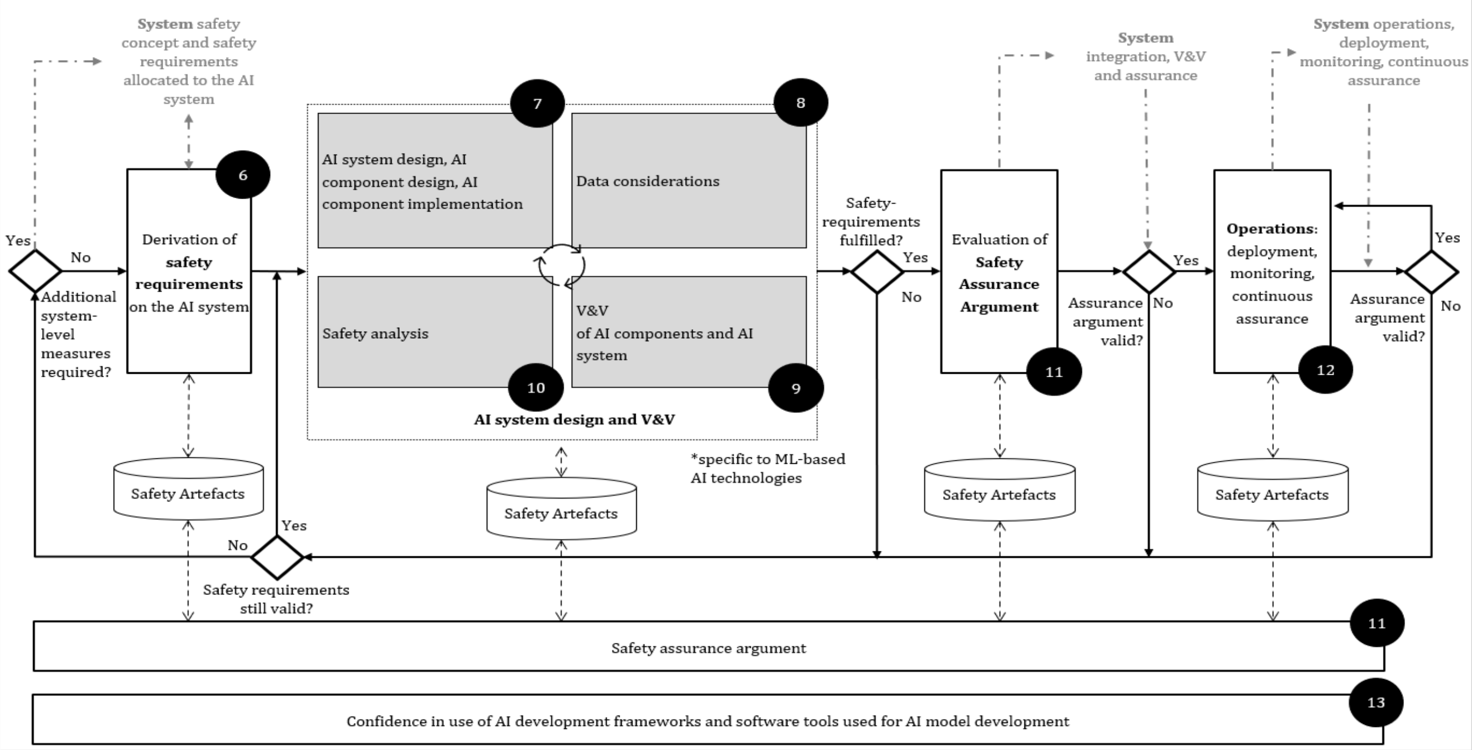
ISO/PAS 8800 – Salient Features

- Supports CI/CD of AI/ML functions (field monitoring and data collection)
- Emphasizes Assurance arguments, besides safety artefacts
- Functional safety-related risks addressed as per ISO/IS 26262
- Performance Limitation risks by extending the concepts and guidance given in ISO/IS 21448
  - Safety requirements are derived by analyzing performance limitations of AI functions.


Contributing error
Can be classified as
Output insufficiency
Caused by
Functional insufficiency
Can be classified as
Performance insufficiency
Insufficiency of specification
contributing systematic error
Caused by
Contributing systematic fault
contributing random hardware error
Caused by
Contributing random hardware fault

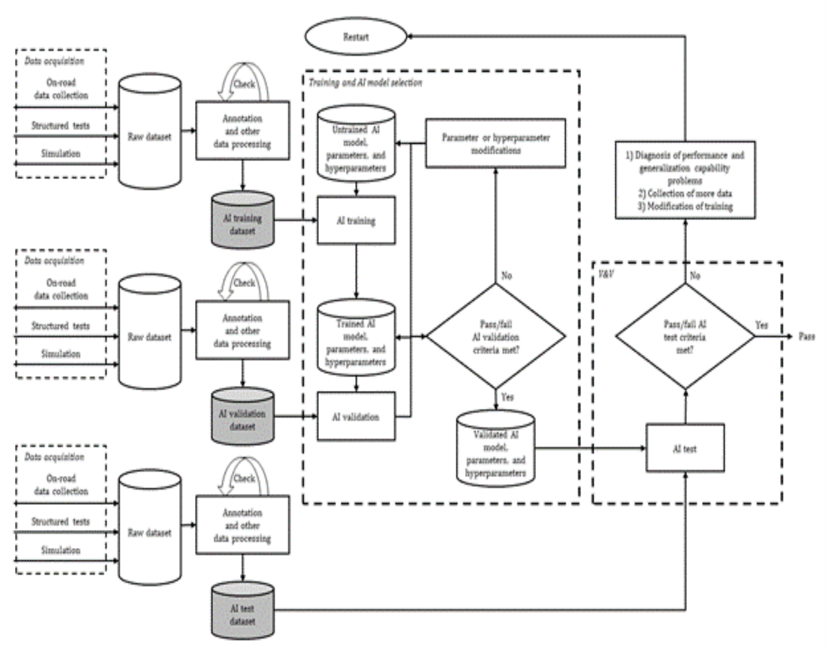
AI Lifecycle

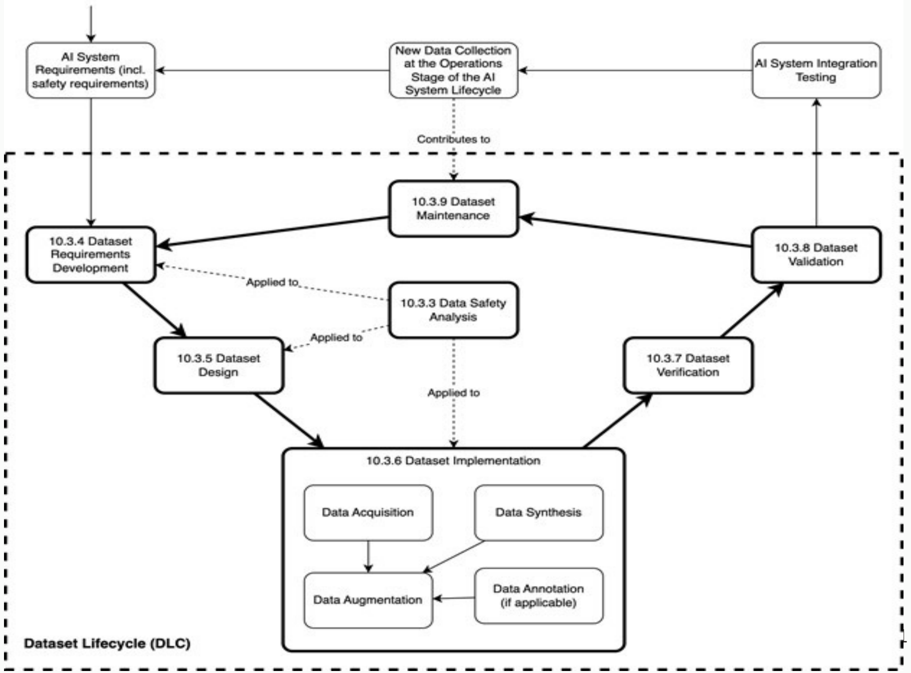
- Exemplary Lifecycle giving rise to Safety Assets forms the basis for the entire document


System safety concept and safety requirements allocated to the AI system
Yes
No
Additional system-level measures required?
6
Derivation of safety requirements on the AI system
7
AI system design, AI component design, AI component implementation
8
Data considerations
Safety analysis
10
V&V of AI components and AI system
9
AI system design and V&V
*specific to ML-based AI technologies
Safety-requirements fulfilled?
Yes
No
Evaluation of Safety Assurance Argument
11
System integration, V&V and assurance
Assurance argument valid?
Yes
No
Operations: deployment, monitoring, continuous assurance
12
System operations, deployment, monitoring, continuous assurance
Assurance argument valid?
Yes
No
Safety Artefacts
Safety Artefacts
Safety Artefacts
Safety Artefacts
Yes
No
Safety requirements still valid?
Safety assurance argument
11
Confidence in use of AI development frameworks and software tools used for AI model development
13

Data Lifecycle

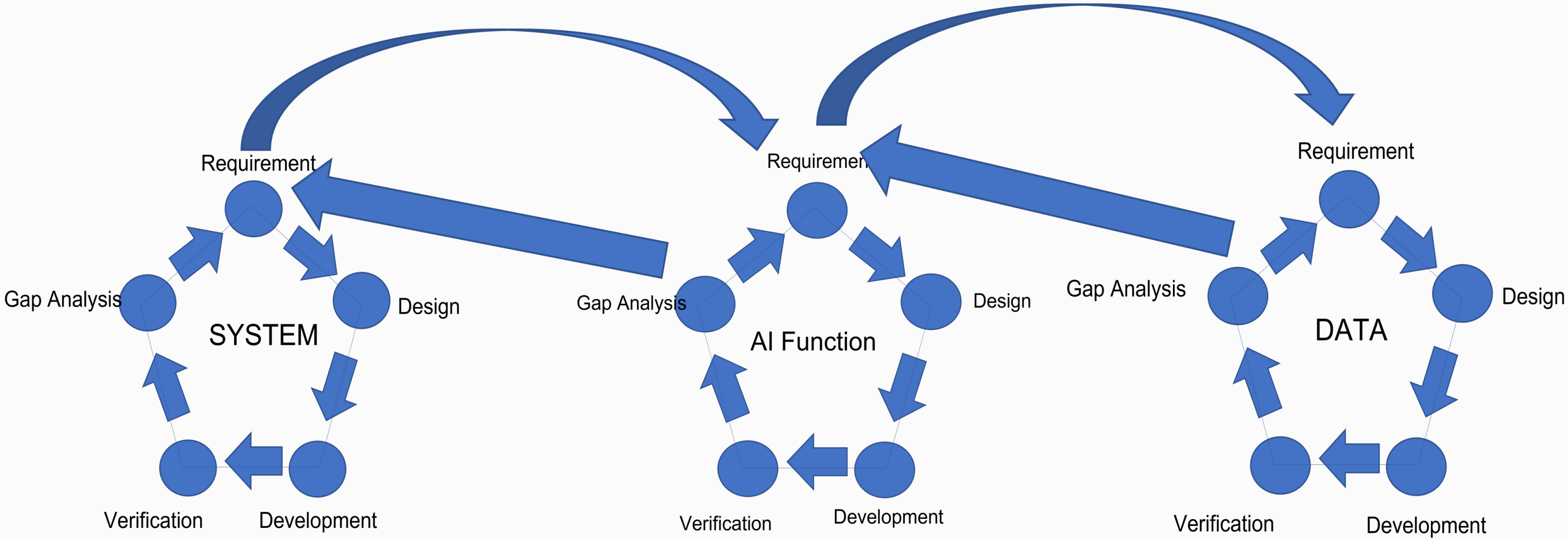
- Data plays a fundamental role in AI system development
  - A dataset lifecycle shall be defined for datasets used in the development of the AI system.
  - The dataset lifecycle shall cover a dataset’s requirements development, design, implementation, verification and validation, safety analysis and maintenance.


Restart
Data acquisition
On-road data collection
Structured tests
Simulation
Raw dataset
Check
Annotation and other data processing
AI training dataset
Data acquisition
On-road data collection
Structured tests
Simulation
Raw dataset
Check
Annotation and other data processing
AI validation dataset
Data acquisition
On-road data collection
Structured tests
Simulation
Raw dataset
Check
Annotation and other data processing
AI test dataset
Training and AI model selection
Untrained AI model, parameters, and hyperparameters
Parameter or hyperparameter modifications
AI training
Trained AI model, parameters, and hyperparameters
AI validation
No
Pass/fail AI validation criteria met?
Yes
Validated AI model, parameters, and hyperparameters
1) Diagnosis of performance and generalization capability problems
2) Collection of more data
3) Modification of training
V&V
No
Pass/fail AI test criteria met?
Yes
Pass
AI test


AI System Requirements (incl. safety requirements)
New Data Collection at the Operations Stage of the AI System Lifecycle
AI System Integration Testing
Contributes to
10.3.9 Dataset Maintenance
10.3.4 Dataset Requirements Development
10.3.8 Dataset Validation
Applied to
10.3.3 Data Safety Analysis
Applied to
10.3.5 Dataset Design
10.3.7 Dataset Verification
Applied to
10.3.6 Dataset Implementation
Data Acquisition
Data Synthesis
Data Augmentation
Data Annotation (if applicable)
Dataset Lifecycle (DLC)

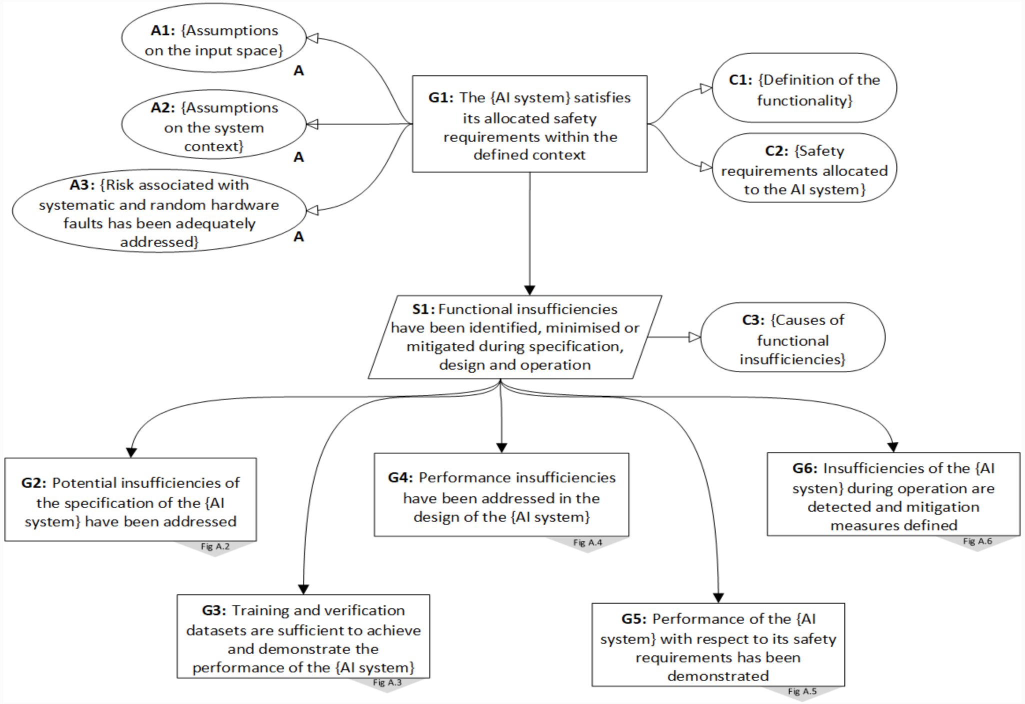
Comprehension of multiple lifecycles

- Traceability from system level to AI and Data life cycle



Assurance arguments

— An assurance argument shall be developed to demonstrate that the AI system fulfils its safety requirements.



A1: {Assumptions on the input space}
A

A2: {Assumptions on the system context}
A

A3: {Risk associated with systematic and random hardware faults has been adequately addressed}
A

G1: The {AI system} satisfies its allocated safety requirements within the defined context

C1: {Definition of the functionality}

C2: {Safety requirements allocated to the AI system}

S1: Functional insufficiencies have been identified, minimised or mitigated during specification, design and operation

C3: {Causes of functional insufficiencies}

G2: Potential insufficiencies of the specification of the {AI system} have been addressed
Fig A.2

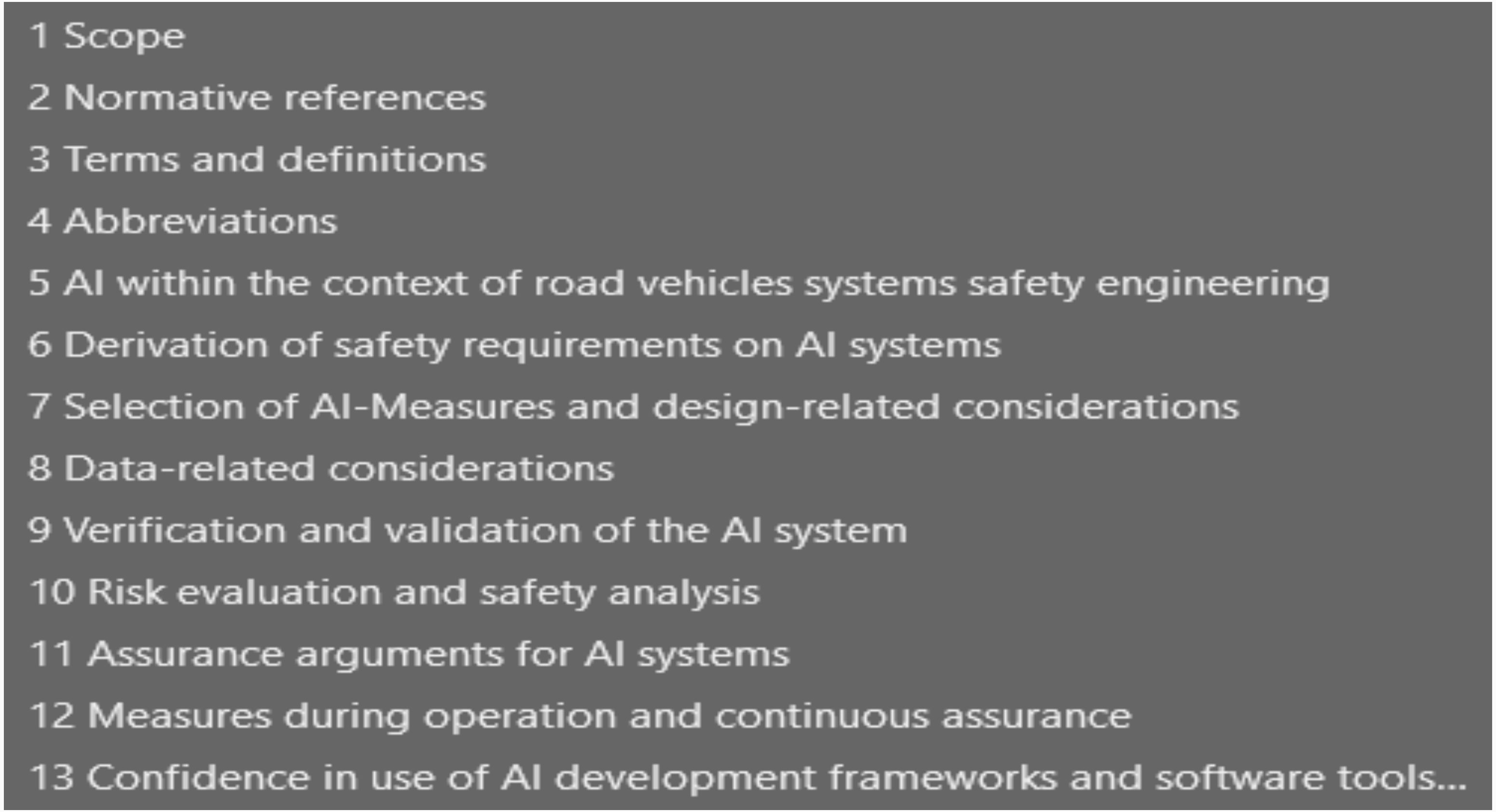
G3: Training and verification datasets are sufficient to achieve and demonstrate the performance of the {AI system}
Fig A.3

G4: Performance insufficiencies have been addressed in the design of the {AI system}
Fig A.4

G5: Performance of the {AI system} with respect to its safety requirements has been demonstrated
Fig A.5

G6: Insufficiencies of the {AI system} during operation are detected and mitigation measures defined
Fig A.6

Chapters - Current Draft



Documentation Timing

- *Working Draft 2* preparation: February and March
- *Working Draft 2* commenting: April and May
- *Working Draft 2* observations and updates: June and July
- *Committee Draft* commenting: August and September
- *Committee Draft* observations and updates: October and November
- Preparation for approval ballot: December
- Approval ballot: January and February ‘24

Questions