DAS SPALTMAß FÜR KI-SYSTEME - WIE SIEHT ES AUS UND WAS SIND AKZEPTABLE GRENZWERTE?

Dr. Rasmus Adler

Program manager “autonomous systems”

Fraunhofer IESE

Übersicht - Das Spaltmaß für KI-Systeme - Wie sieht es aus und was sind akzeptable Grenzwerte?

- Motivation
- Was ist ein KI-System?
- Welche Qualitätseigenschaften mit welcher Qualität messen?

Motivation: „AI made in Germany“

Die Bundesregierung

Nationale Strategie für Künstliche Intelligenz
AI Made in Germany

- KI aus Deutschland soll zum Gütesiegel werden
- Sich strategisch auf Qualität und Prüfung zu fokussieren ist naheliegend
  - Qualität ist eine Stärke im Maschinen- und Karosseriebau
    - Das deutsche „Spaltmaß“ hat symbolischen Charakter
  - Die TIC-Industrie ist in Europa und Deutschland sehr stark
  - Mit USA und China mitzuhalten bezüglich Investitionen und Zugriff auf Daten ist schwer
  - Vorbildhafte Regulatorik und Normierung bieten eine Chance für die TIC-Industrie

[1]

- Der erste Schritt für Regulierung und Normierung ist eine Definition von dem was reguliert werden soll

[1] https://www.ki-strategie-deutschland.de/home.html

Fraunhofer IESE

Ein „KI – System“ ist laut dem EU AI Act eine Art von Software

- Klare Begrenzung auf Software aber kaum Einschränkung durch
  - „listed in Annex 1“
    - kann breit interpretiert werden
      - Beispiel: „Statistical approaches“
  - „…for a given set of human defined objectives…“
  - „…outputs such as content, …“

- Grundlegende regulatorische Anforderungen sollen durch technische Normen verfeinert werden (New Approach, New Legislative Framework)
  - Beispiel: Maschinenrichtlinie und deren harmonisierte Normen
- -> Die Definition in Gesetzgebung und Normung sollten konsistent sein

AI system in EU AI Act

artificial intelligence system’ (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with

ANNEX I List
(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
(c) Statistical approaches, Bayesian estimation, search and optimization methods.

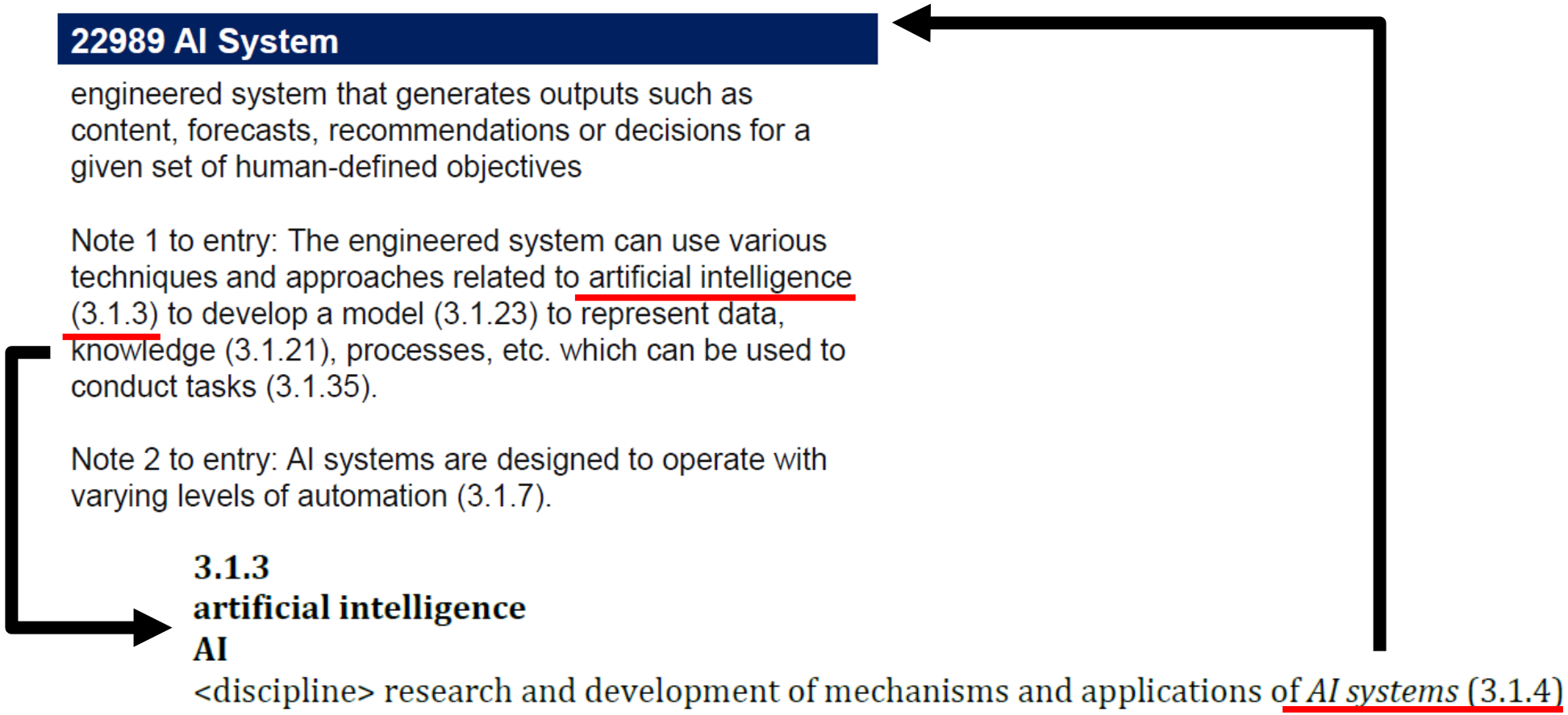
Ein „KI – System“ ist nach ISO/IEC 22989 ein „engineered“ System

- Ein „engineered“ System kann deutlich mehr sein als Software!

22989 AI System	AI system in EU AI Act
engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives Note 1 to entry: The engineered system can use various techniques and approaches related to artificial intelligence (3.1.3) to develop a model (3.1.23) to represent data, knowledge (3.1.21), processes, etc. which can be used to conduct tasks (3.1.35). Note 2 to entry: AI systems are designed to operate with varying levels of automation (3.1.7).	artificial intelligence system’ (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with ANNEX I List (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods.

Ein „KI – System“ ist nach ISO/IEC 22989 zyklisch definiert

- Zyklen sollten in Definitionen vermieden werden



Ein „KI – System“ ist nach ISO/IEC 22989 ein „autonomes“ System

- Die einzige Einschränkung zu „engineered system“ kommt durch Note 2!

22989 AI System

engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives

Note 1 to entry: The engineered system can use various techniques and approaches related to artificial intelligence (3.1.3) to develop a model (3.1.23) to represent data, knowledge (3.1.21), processes, etc. which can be used to conduct tasks (3.1.35).

Note 2 to entry: AI systems are designed to operate with varying levels of automation (3.1.7).

Jedes System generiert Ausgaben (Energie, Masse, Information)!

Das ist keine Bedingung!

Warum müssen die Level variieren?

Was bedeuten die Level?

Ist ein autonomes KI-System per Definition fehlerhaft?

- Das höchste Level bezieht sich auf die Einsatzumgebung „**des“** Systems und Ziele „**des“** Systems
- Aber in der Definition von KI-Systems heißt es „…for a given set of **human-defined** objectives“

Level of automation	Comments
Autonomy	The system is capable of modifying its operating domain or its goals without external intervention, control or oversight.

- Wurden die Ziele und die Einsatzumgebung von einem Menschen definiert?
  - Wenn ja, welcher Mensch? Der Entwickler?
  - Wenn ja, dann ist autonomes Verhalten per Definition [1] fehlerhaft !?!

- Der Rest der Terminologie zu „Level of Automation“ wirft weitere Fragen auf

[1] A. Avizienis, J. . -C. Laprie, B. Randell and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," in IEEE Transactions on Dependable and Secure Computing, vol. 1, no. 1, pp. 11-33, Jan.-March 2004, doi: 10.1109/TDSC.2004.2.

Ist ein autonomes System ein Spezialfall von einem automatisiertem System?

		Level of automation
Automated system	Autonomous	Autonomy
	Heteronomous	Full automation
		High automation
		Conditional automation
		Partial automation
		Assistance
		No automation

Kann man Systeme wirklich automatisieren oder automatisiert man eher Aufgaben / Prozesse etc. durch den Einsatz technischer Systeme?

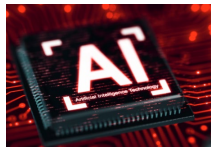
Die Anwendungsregel VDE-AR-E 2842-61 „Entwicklung und Vertrauenswürdigkeit von autonom/kognitiven Systemen“ definiert KI-Element und KI-Technologie

Socio-technical system

Autonomous cognitive System

E/E-Architecture

ECU



AI HW SW

Initiation

Solution Concept

Trustworthiness Analysis

Autonomous/Cognitive System (or System of Interest)

Technological Element

AI-Development

AI-Blueprint

Other Technology Development

IV&V of the System

IV&V of the Solution

Acceptance and Release

back to appropriate phase of the lifecycle

After Release of the Solution

Part 6

Part 5 Development at Technology Level

Part 4 Development at System Level

Part 3 Development at Solution Level

Übersicht - Das Spaltmaß für KI-Systeme - Wie sieht es aus und was sind akzeptable Grenzwerte?

- Motivation
- Was ist ein KI-System?
- **Welche Qualitätseigenschaften mit welcher Qualität messen?**

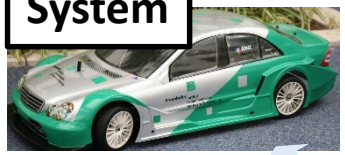
Qualitätsaspekte der VDE-AR-E 2842-61

Engineering der Qualitätsaspekte über alle Ebenen hinweg (Beispiel Safety)

Socio-technical system (of systems)

"AI system?"

System

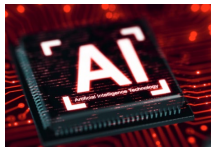


"AI system?"

E/E-Architecture

ECU

Technology



"AI system!" HW SW

SIL?

SIL!

VDE-AR-E 2842-61

System safety → MIL 882E,…

Initiation

back to appropriate phase of the lifecycle

Acceptance and Release → After Release of the Solution

Solution Concept

IV&V of the Solution

Product / machinery safety → DIN EN ISO 12100,…

System (or System of Interest)

IV&V of the System

Functional safety → IEC 61508,…

Technological Element

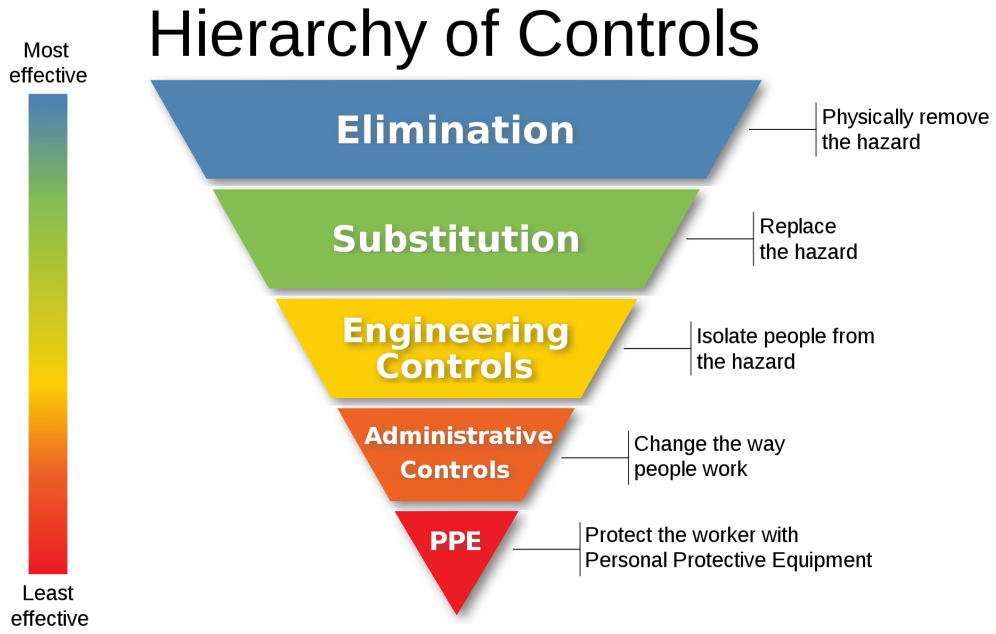
AI-Development

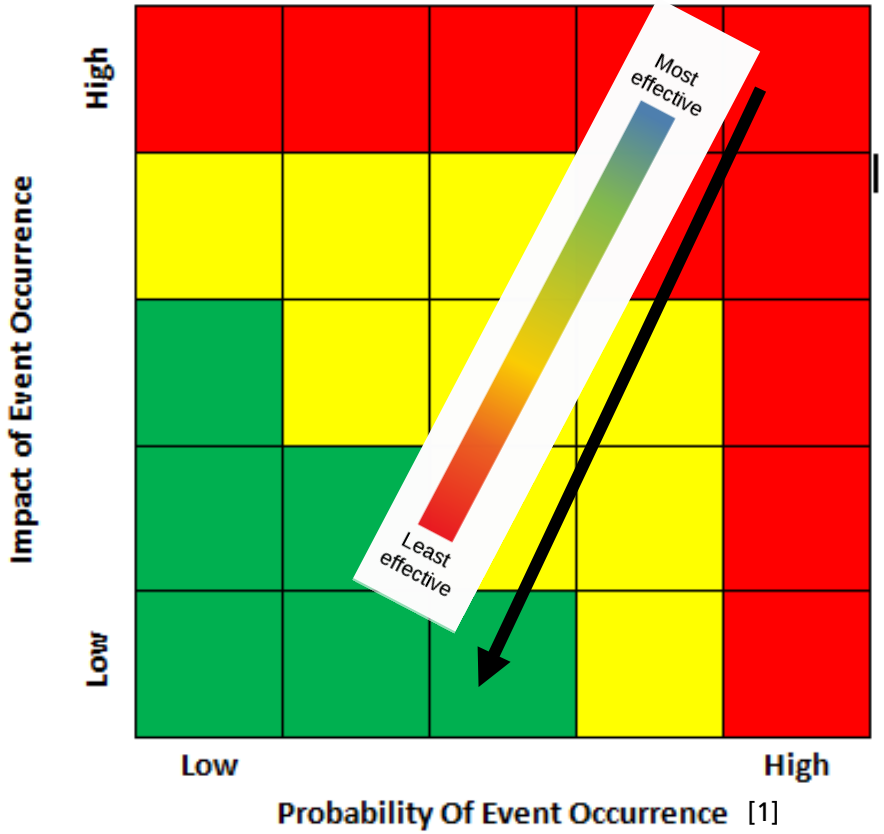
AI-Blueprint

Other Technology Development

Fraunhofer IESE

Es gibt eine Hierarchie von Maßnahmen



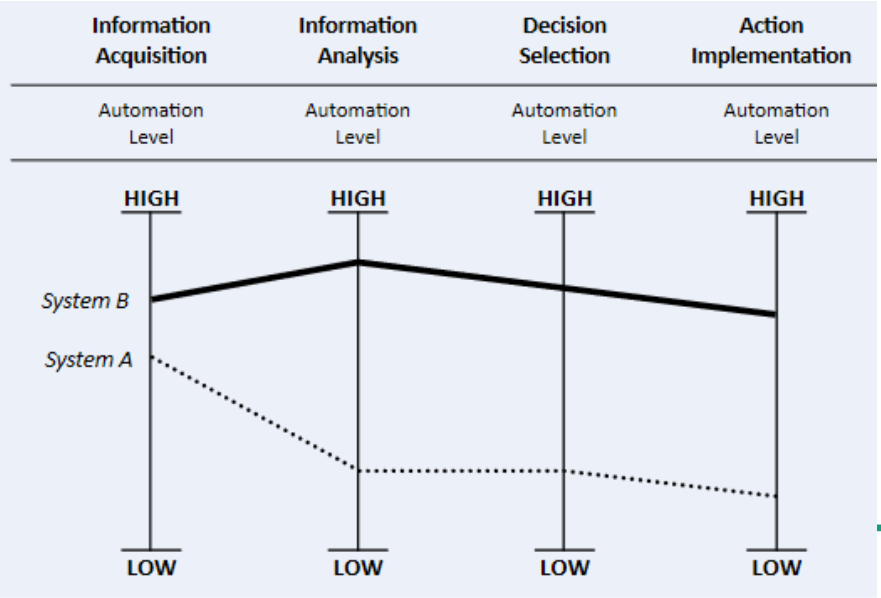


[1]

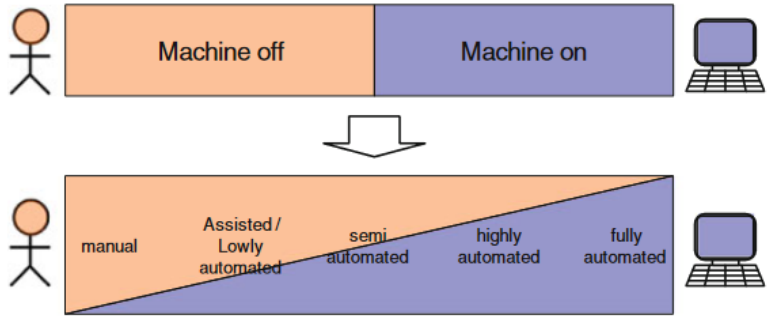
- Technische Systeme können in verschiedenen Ebenen vorkommen
- Die Funktion dieser Systeme kann mit SW / HW / KI realisiert werden
- -> Funktionale Sicherheit fokussiert auf die korrekte Funktion dieser Systeme

Maßnahmen auf der obersten Ebene: Allokation von Funktionen und Mensch-Maschine Interaktion

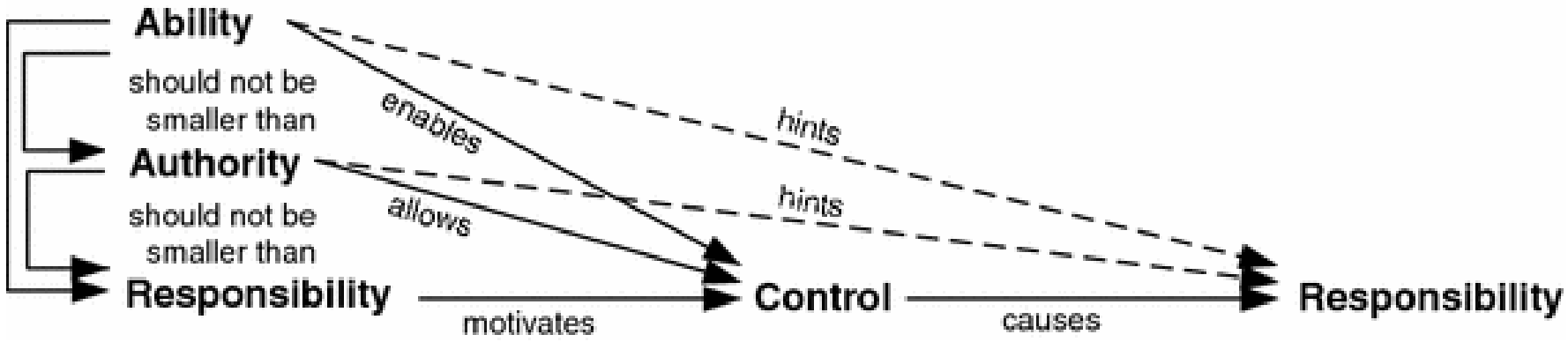
- Automatisierung muss im soziotechnischen Kontext des Gesamtsystems betrachtet werden (-> Solution Level)
- Automatisierung ändert fast immer nur die Rolle von Menschen (Beispiel: Safety-Rolle bei Betrieb und Wartung)
- Berücksichtigung von „Typen“ und „Leveln“ der Automatisierung



Maßnahmen auf der obersten Ebene: Allokation von Funktionen und Mensch-Maschine Interaktion

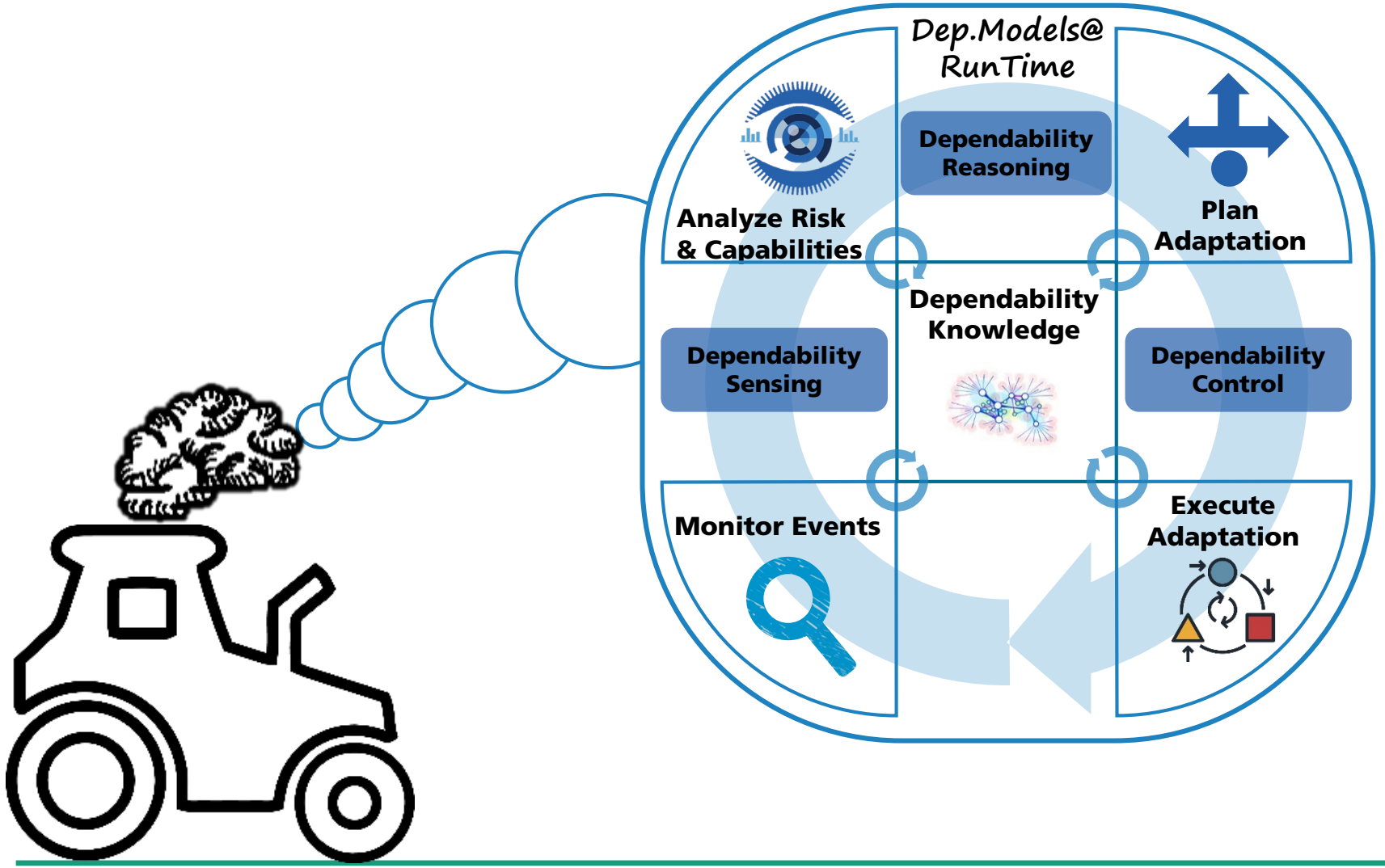


- Angemessene Allokation im Hinblick auf „Fähigkeiten“, „Befugnis“ und „Verantwortung“
  - Mehr Verantwortung als Befugnis zu haben ist nicht sinnvoll
    - Beispiel: Verantwortung des Operators im „Autonomen Modus“
  - Mehr Befugnis als Fähigkeit zu haben ist nicht sinnvoll
    - Beispiel: Automatisierungsmüdigkeit



Flemisch, F., Heesen, M., Hesse, T. *et al.* Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. *Cogn Tech Work* **14**, 3–18 (2012). https://doi.org/10.1007/s10111-011-0191-6

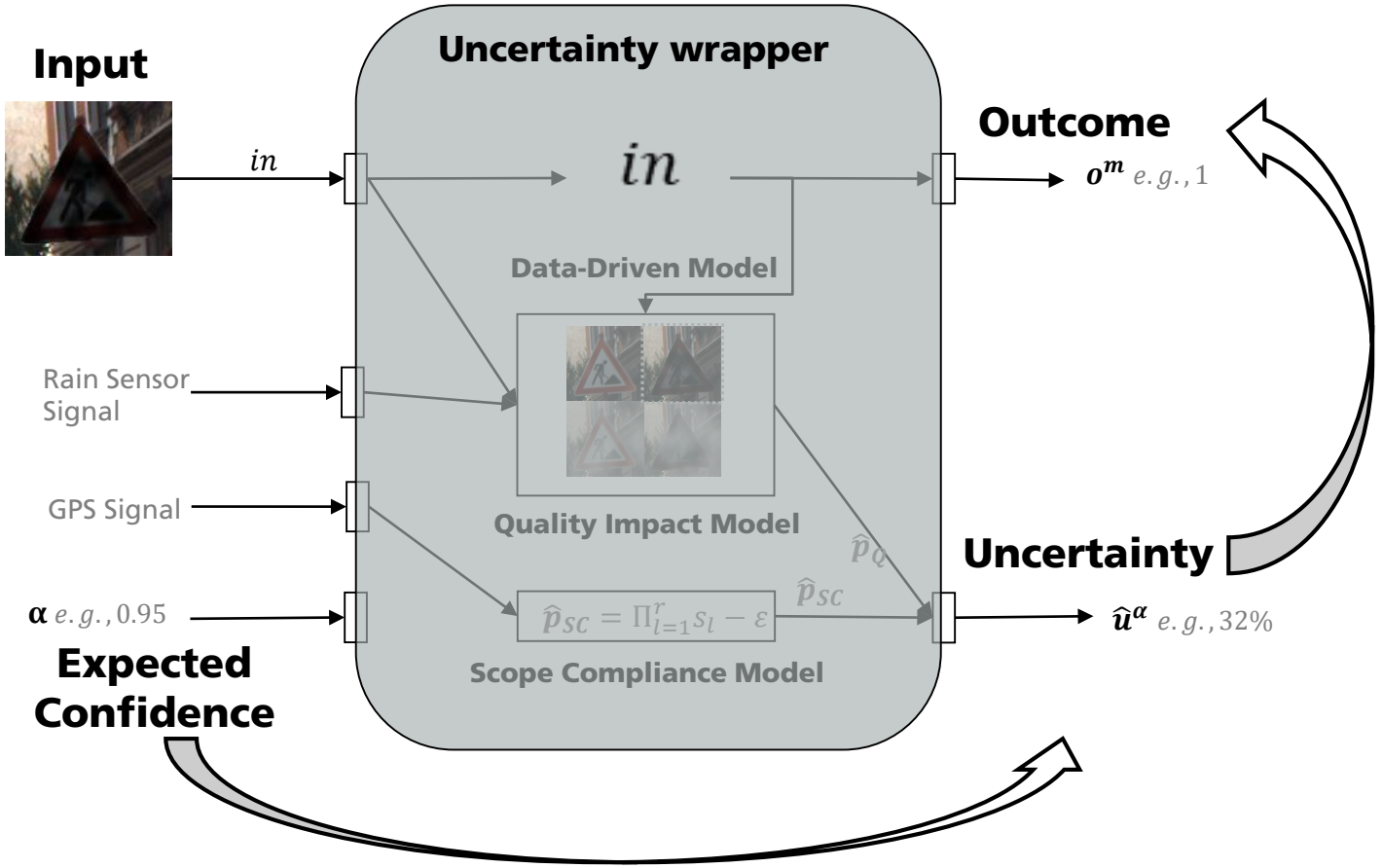
Maßnahme auf Systemebene z.B. „Laufzeit Risikomanager“



Related to **ISO 21815-1:2022**

Earth-moving machinery — Collision warning and avoidance

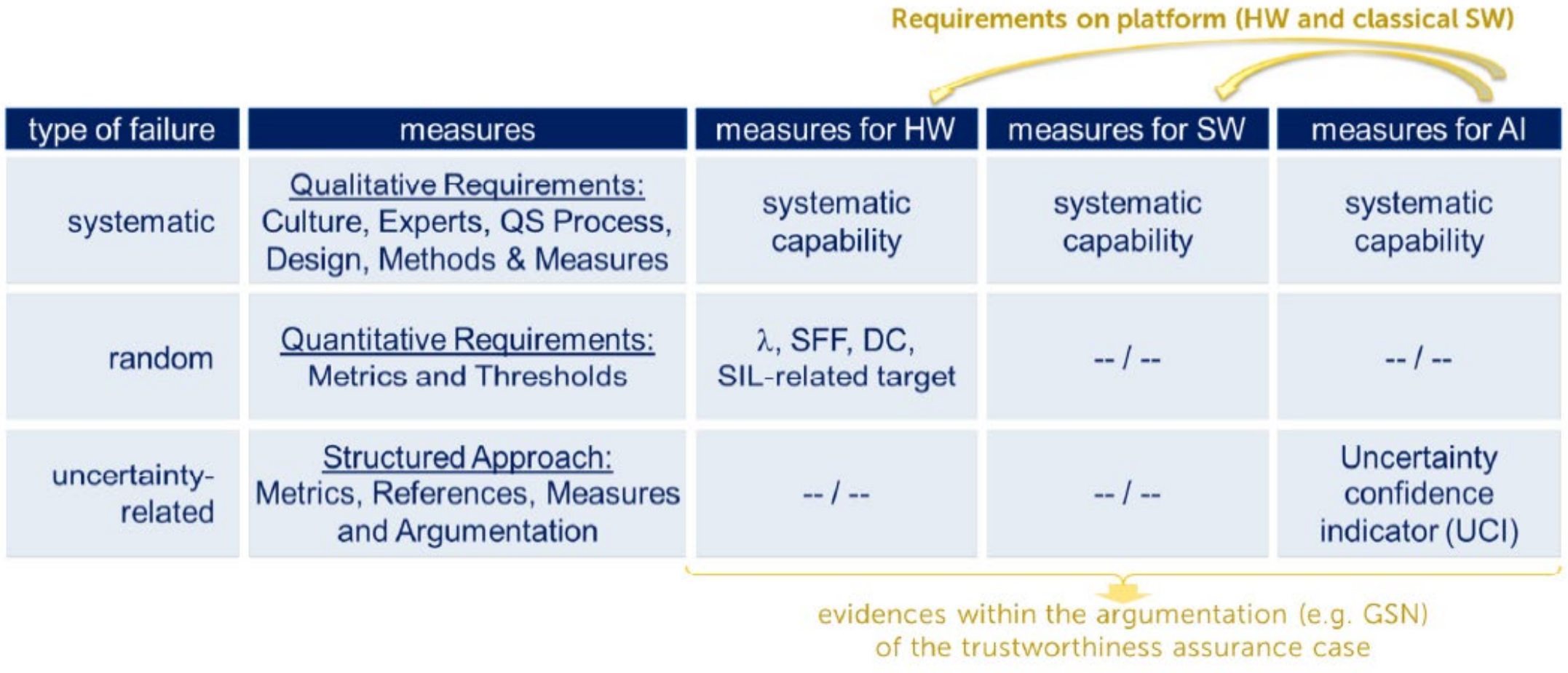
Maßnahme auf Technischer Ebene: Unsicherheiten messen und behandeln



Wahrscheinlichkeit, dass die Ausgabe falsch ist aufgrund von Unsicherheiten durch (1) inhärente Limitierungen des gelernten Modells, (2) limitierte Eingabequalität während der Anwendung (Regen, etc.), und (3) Abweichungen zwischen modelliertem Kontext und Anwendungskontext

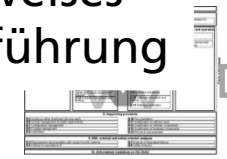
Wahrscheinlichkeit, dass die Fehlerwahrscheinlichkeit zu niedrig ist

Maßnahme auf Technischer Ebene: Unsicherheiten messen und behandeln

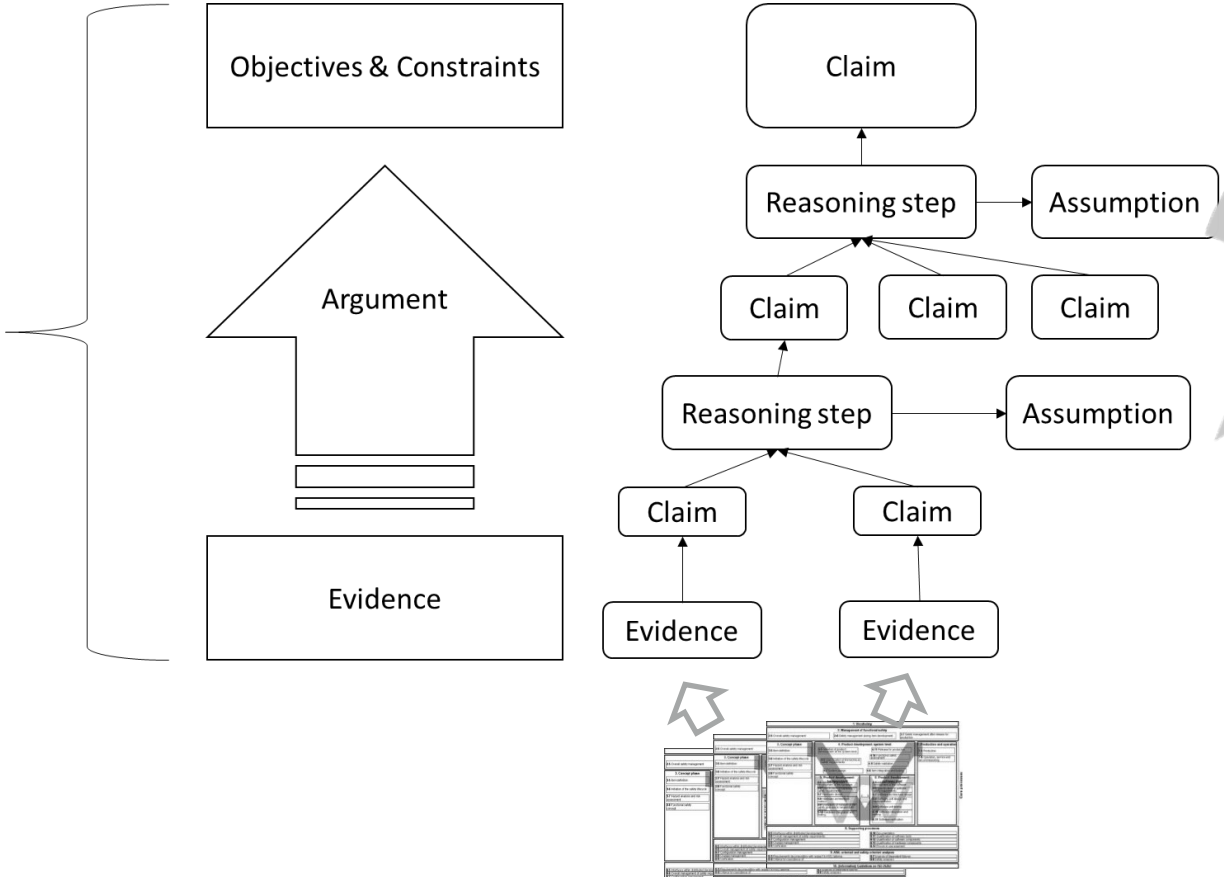


Im Assurance Case erklären warum die Maßnahmen ausreichen

Anwendungsregeln für die Erstellung eines Nachweises vor Markteinführung



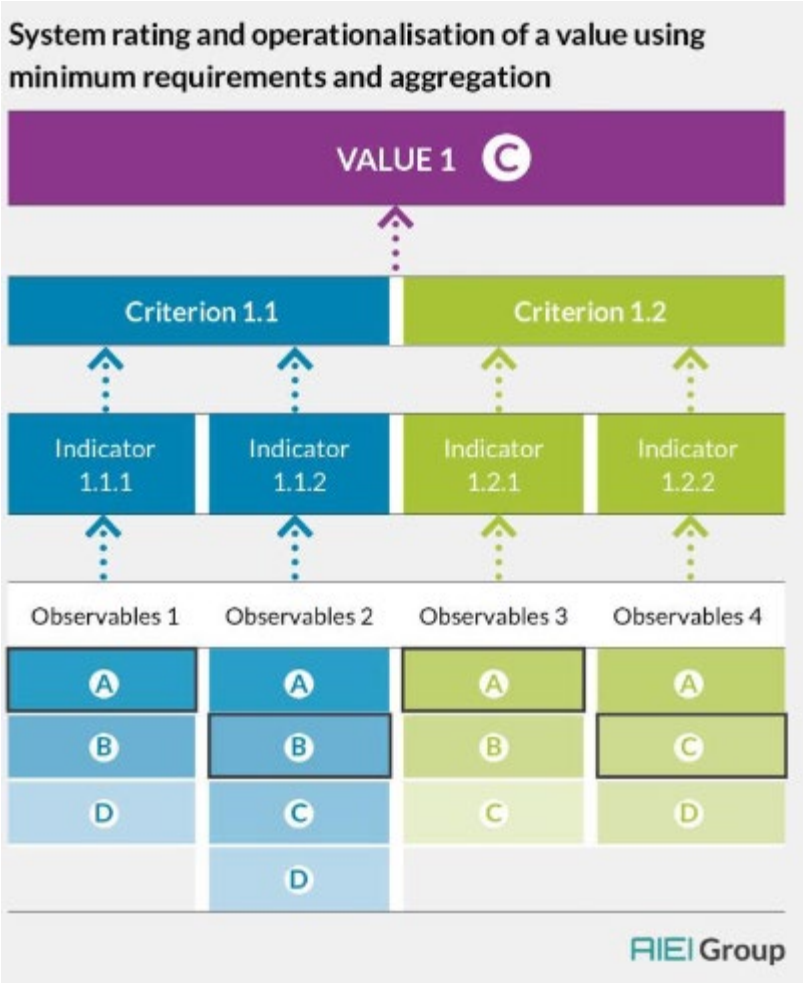
Assurance case

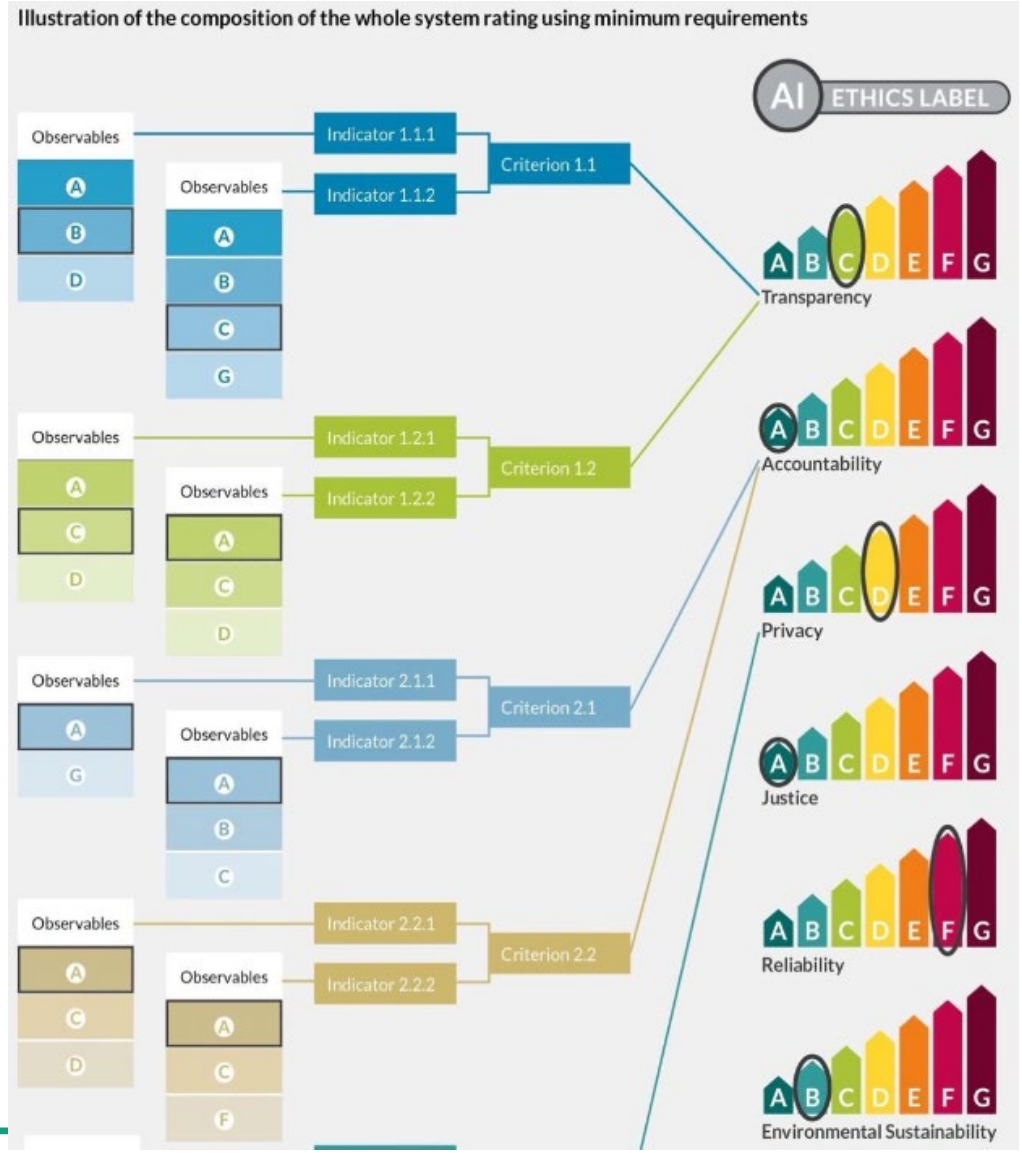


Anwendungsregeln für die „Wartung“ und „Verbesserung“ eines Nachweises nach Markteinführung

Nachweise: Arbeitsprodukte, die bei der Anwendung von Normen entstehen, wie Gefahren- & Risikoanalyse, Testreports,…

Das VCIO Modell als leichtgewichtige Variante eines Assurance Cases





https://www.vde.com/de/presse/pressemitteilungen/ki-ethik-messbar-machen

Fraunhofer IESE

Zusammenfassung

- Es gibt noch keinen (guten) Konsens darüber was ein KI-System ist
- Es gibt Konsens darüber wie man grundsätzlich Qualität misst
  - Die Ansätze „Goal Question Metric“, „VCIO“, „Assurance Cases“ harmonisieren
- Die Ausgestaltung der Ansätze um harmonisierte Normen für den AI Act zu entwickeln wirft noch viele Fragen auf
  - Trade-off zwischen flexibler / zielbasierter Norm und eindeutiger / regelbasierter Norm
  - Die Qualität des Qualitätsmaß festlegen
    - Beispiel Assurance Cases
      - Wie lässt sich die induktive Stärke eines Arguments objektiv bewerten?

Thank you for your interest

Kontakt:

Dr. Rasmus Adler
Rasmus.adler@iese.fraunhofer.de
Fraunhofer IESE
Programm Manager Autonome Systeme