Bundesamt
für Sicherheit in der
Informationstechnik

Deutschland
Digital•Sicher•BSI•

TAISEC & TAISEM
An approach to horizontal Criteria for AI Systems Evaluation & Certification

Thomas Barz, Daniel Loevenich [Federal Office for Information Security]

Dr. Maximilian Poretschkin [Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS]

October 2023

Agenda

Horizontal Trustworthy AI Conformity Assessment

AI Trustworthyness Functionality Testing

Effectiveness of [Counter-] measures

Correctness and Assurance

A uniform Framework for AI Conformity Assessment

Horizontal Trustworthy AI Conformity Assessment: Mission

To operationalize the recommendations for action of the Standardization Roadmap AI that concern the technical requirements for AI systems, it is necessary to propose a national implementation programme.

The mission of this implementation programme is to develop such testing and quality assurance standards as central technical components of the action framework in a timely and needs-based manner, and to enable them to be updated in the future on the basis of economic and technical progress.

Uniform Evaluation & Certification Framework for Trustworthy AI Applications

Horizontal Trustworthy AI Conformity Assessment: Strategic Objectives

Trustworthiness of the **entire Supply Chain** becomes transparent with **Conformity Assessment**

The necessary **evaluation criteria and test procedures** are to be developed

This evaluation bases shall be **applicable to hybrid solutions** as well as **embedded components**

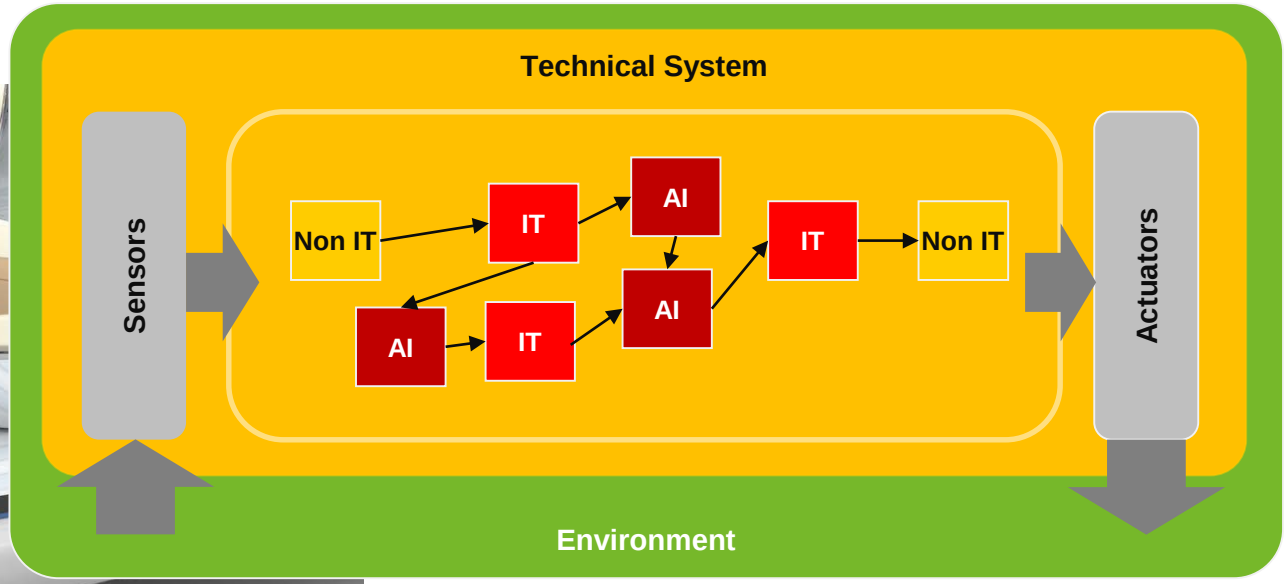
EU-AIA: The evaluation bases serve as foundation for a **horizontal AI standard**

An **easy market access for SMEs** with acceptable costs will be facilitated

Vertical and sectoral standards should be based a horizontal standard for trustworthy AI

Uniform Evaluation & Certification Framework for Trustworthy AI Applications

Functionality of embedded AI Applications



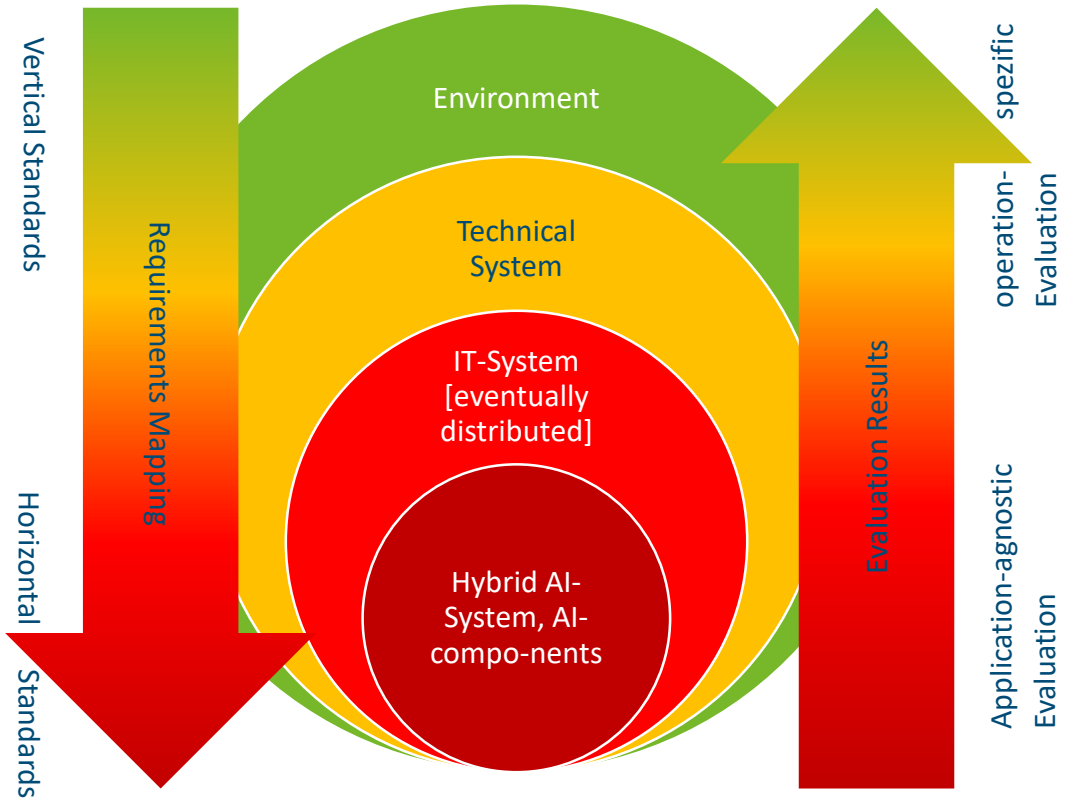
From Application Risks to Specs of AI-Trustworthyness Functions

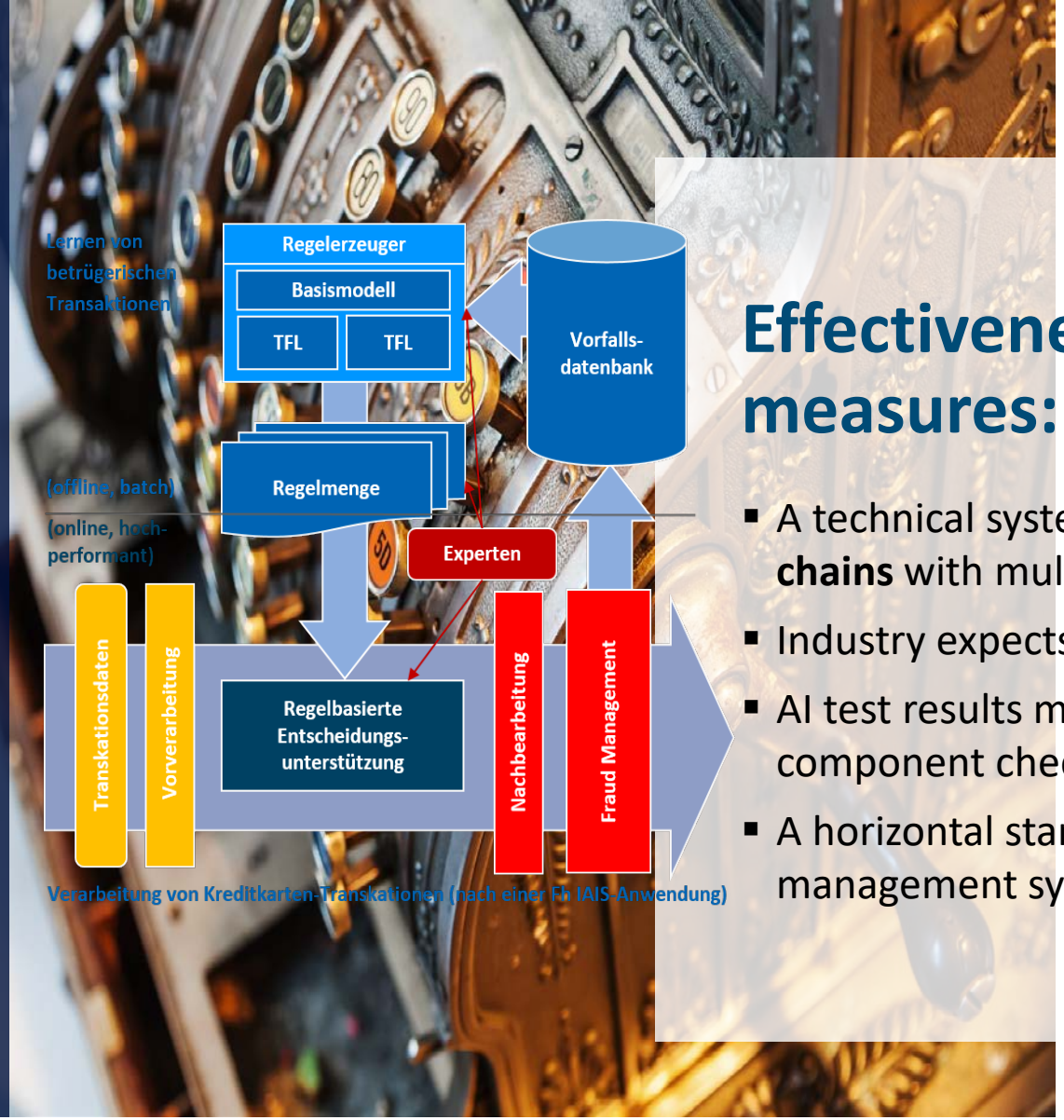
- **Conformity Assessment** considers the complex technical system (e.g. a vehicle) as a whole entity in the application context
- **Risks and requirements for the AI-components** have to be derived from this
- Such **Operationalizations** are essentially **mappings** into AI requirements
- This results in evaluation requirements for each component of usually hybid AI solutions (**application-agnostic** perspective)

Functionality Testing and Evaluation

Horizontal and Vertical Standards Concept:

- **Downgrading Risks** to AI Components
- **Conformity Testing and Evaluation** of AI Components (all CA types)
- Upgrading and **Composition** of Evaluation Results
- **Uniform Conformity Assessment Framework**
- Application of schemes and **introduction to markets worldwide**



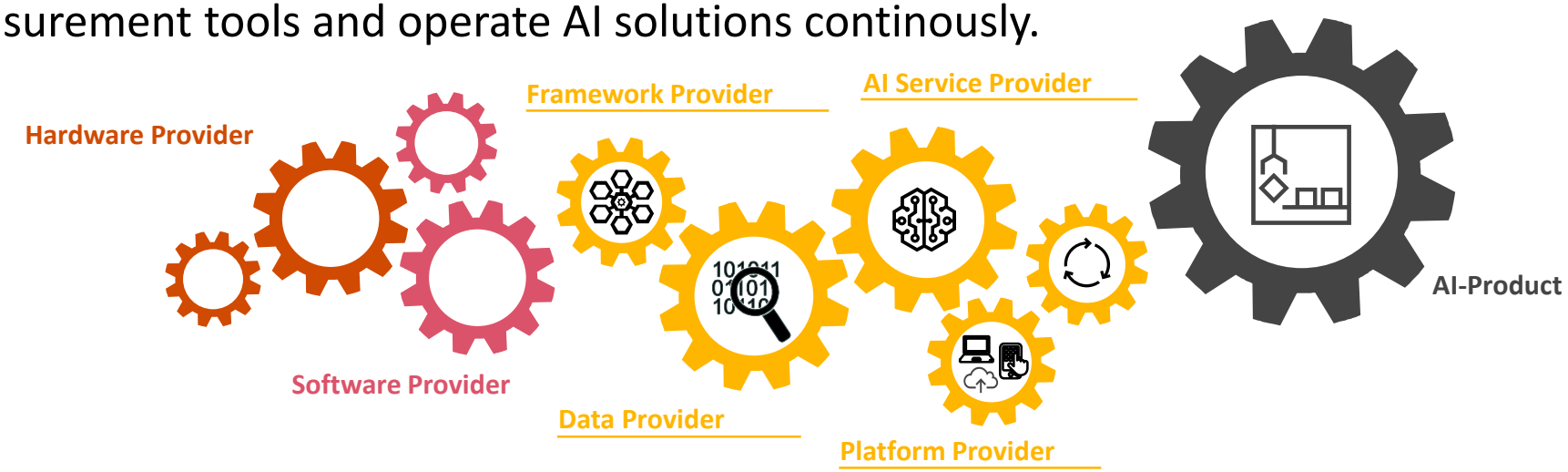


Effectiveness of [Counter-] measures: Principles

- A technical system usually follows **existing test regulations for supply chains** with multiple actors, e.g. OEMs
- Industry expects **cascadable evaluations for (distributed) AI-systems**
- AI test results must be **composable** (for instance in case of individual component checks of hybrid systems)
- A horizontal standard for trustworthy AI must be referable for AI management systems (**AIMS**)

Effectiveness of [Counter-] measures: Processes

AI technologies are used within the compete AI supply chain. They offer full service enterprise customer support with AI experts for development, IDEs, Frameworks and quality measurement tools and operate AI solutions continously.



A project for an horizontal standard for trustworthy AI shall evaluate these resources.

Correctness & Assurance: Conformity Assessment Methods

Selection = Selection of applicable requirements, choice of methods, planning, sampling

Determination = Activities to collect evidence of conformity with regard to the specified requirements, i.e. analyses, tests, evaluations, investigations, audits, tests, inspections, validations, verifications, etc.

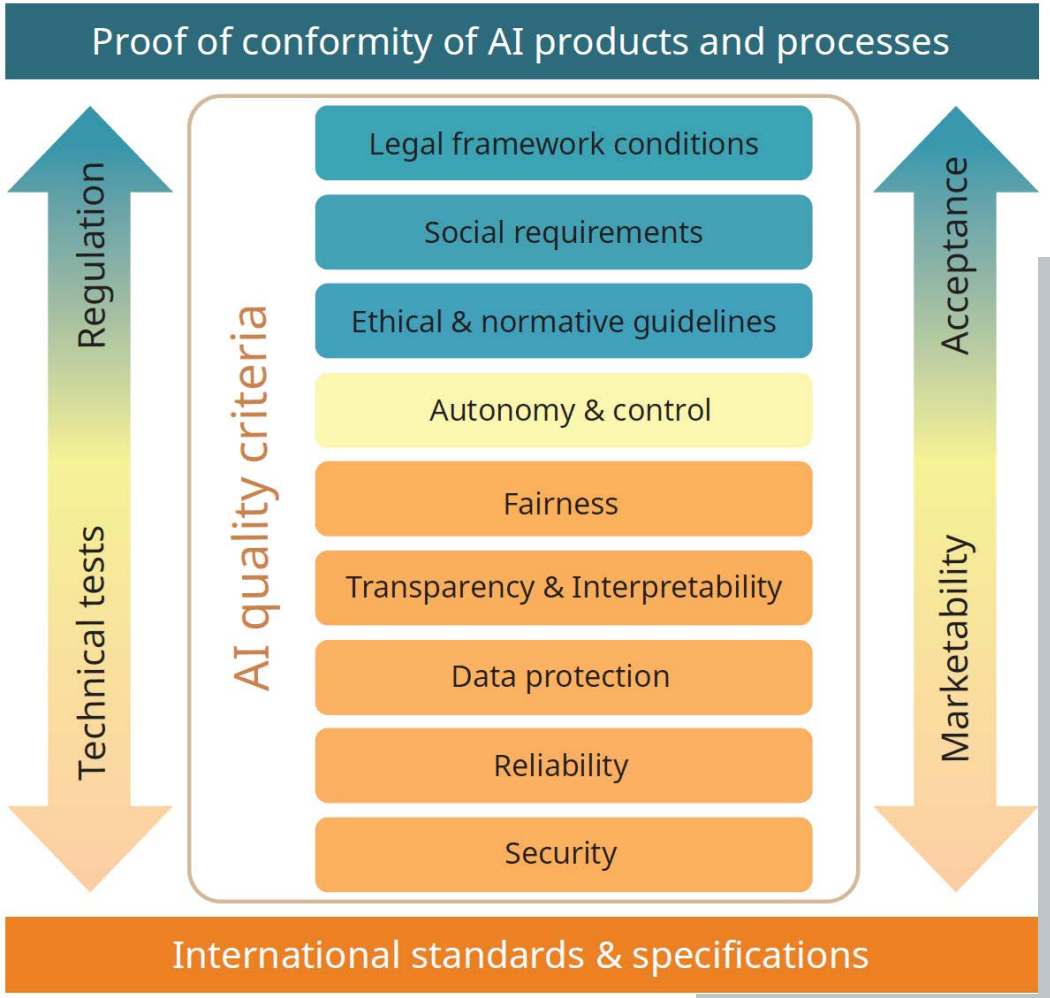
Review = Conclusion regarding suitability, adequacy and the sufficient amount of evidence collected

Decision = Deciding whether or not the assessed object has been shown to conform to the specified requirements

Attestation = Formal issue of the statement of conformity, e.g. test report (test passed/failed) or certificates

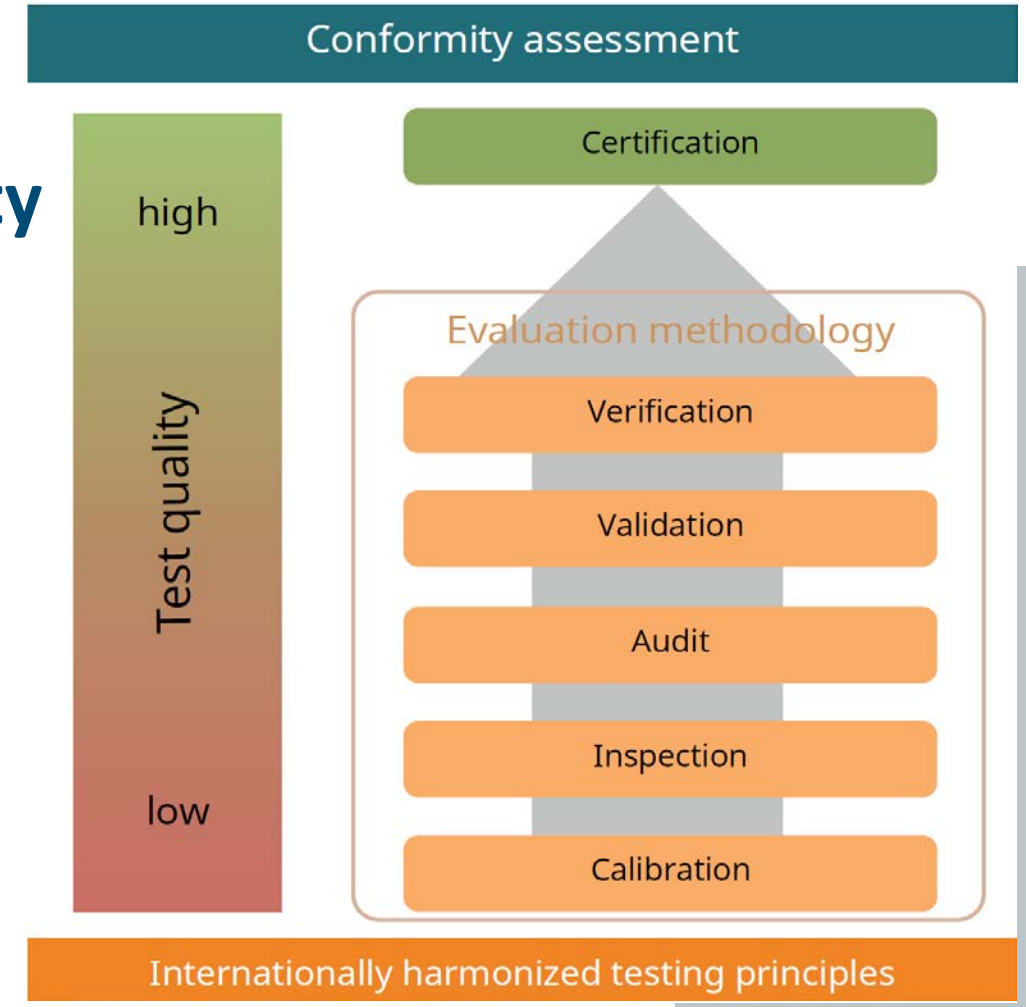
Correctness & Assurance: Categories of AI quality criteria

- Audit
  - Check that an organization's processes, practices and procedures meet certain requirements formulated in a standard. This check is usually based on a list of criteria derived from the underlying standard.
- Validation
  - Confirmation of the plausibility of a specific use or application purpose by providing objective evidence that specified requirements have been met.
- Verification
  - Confirmation of truthfulness by providing objective evidence that specified requirements have been met.



Correctness & Assurance: Evaluation methodology and test quality

- Evaluation based upon **predefined Assurance Levels**
  - Testing, inspection and validation/verification activities may be performed by the supplier (first party) of the object to be evaluated or by a person/organization with an interest as a user of that object (second party).
- Certification
  - Confirmation by a third party relating to an object of conformity assessment (accreditation excluded). A “third party” is independent of the supplier of the object of the conformity assessment activity and has no interest as a user. Certifications are only offered by independent bodies.
- Test & Evaluation Facilities



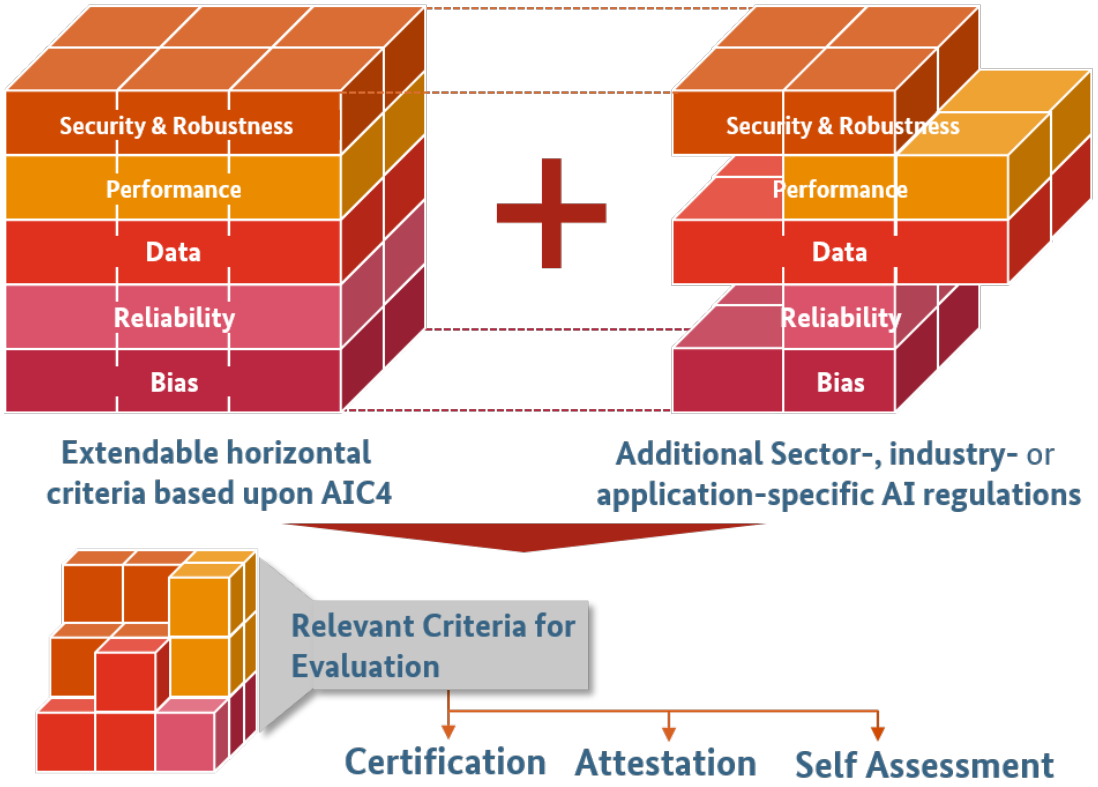
Correctness & Assurance: Use Case Projects Outline

	Phase 1:	Phase 2:	Phase 3:	Phase 4:
Focus:	Scoping, Risk Analysis & TOE	Evaluation Procedure Development	Procedure Execution	Tool and Framework Validation
Outcome (Documents):	Mapping System Description Minutes	Procedure Description Protocol	Evaluation Documents (dependable on the conformity assessment type) Evaluation Report	Evaluation Documents Evaluation Report
Format:	Workshops	Coordination Conference with unanimous vote	Depends on the type of conformity assessment	Certification Report
	Mandatory: Generalization of Requirements up to Criteria and Evaluation Process Definition		**Optional: Pilot Evaluation, Procedure Validation, Assurance Methods Assessment (Levels)**	

Uniform Evaluation & Certification Framework for Trustworthy AI Applications: Deliverables

TAISEC: An extendable set of **horizontal, application agnostic criteria** (“**Trustworthy AI Systems Evaluation Criteria**”)

TAISEM: An extendable set of valid **AI evaluation procedures** (“**Trustworthy AI Systems Evaluation Methodology**”), applicable to all three types of conformity assessment (self assessment, attestation and certification),



Uniform Evaluation & Certification Framework for Trustworthy AI Applications: Deliverables

A proposal for an **application procedure** to extend the methodology and to implement it within the ongoing standardization process,

A proposal for an **application procedure** to extend this criteria and to implement the procedure within the ongoing standardization process,

A **procedure for mappings** of vertical application specific requirements into horizontal criteria requirements,

The **Guidance documents for production, application, and support** for all parties involved in the corresponding AI Eco-Systems. In the context, the framework establishes additional guidelines on how to integrate evaluation activities into an AIMS.

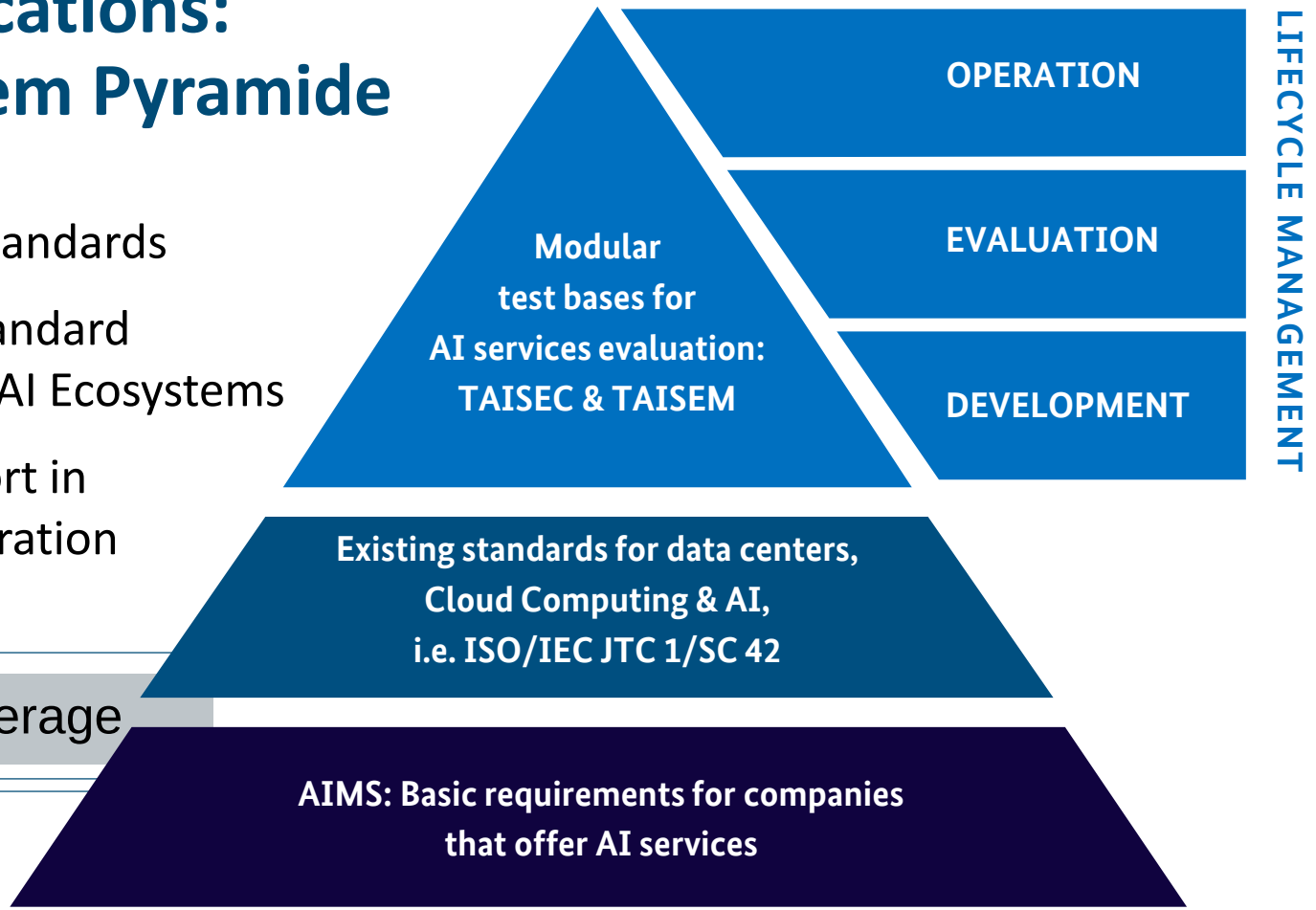
Uniform Evaluation & Certification Framework for Trustworthy AI Applications: Standardization System Pyramide

Transfer of AI Trustworthiness Standards

Conformity Assessment of AI-Standard Solutions lead to Acceptance of AI Ecosystems

Worldwide End Customer Support in regulated sectors: Market Penetration

Complete AI Ecosystem Coverage



Uniform Evaluation & Certification Framework for Trustworthy AI Applications: Project Roadmap

	Definition of evaluation bases	Applicability and Market Penetration	Publication Phase	International Standardization
	Q1 - 2024	Q3 - 2024	Q1 2025	2025 →
	Development of **criteria, methodology, scheme** with use cases	**Validation** and Acceptance on basis of relevant Use Cases	Publication within **three Workshops -** Europe, USA, Asia	Transfer and Harmonization in **hEN/ISO Standard**

AI Standardization Programme:

Summary of Lighthouse Projects

- **Horizontal Standardization Projects:**
  - **TAISEC – Trustworthy AI Systems Evaluation Criteria**
  - **TAISEM – Trustworthy AI Systems Evaluation Methodology**

Do you want to join?

- Use cases in various sectors:
  - Health Care
  - Financial Services
  - Agriculture Devices
  - Critical Infrastructures

Thank you! Q&A?

Deutschland
Digital•Sicher•BSI•

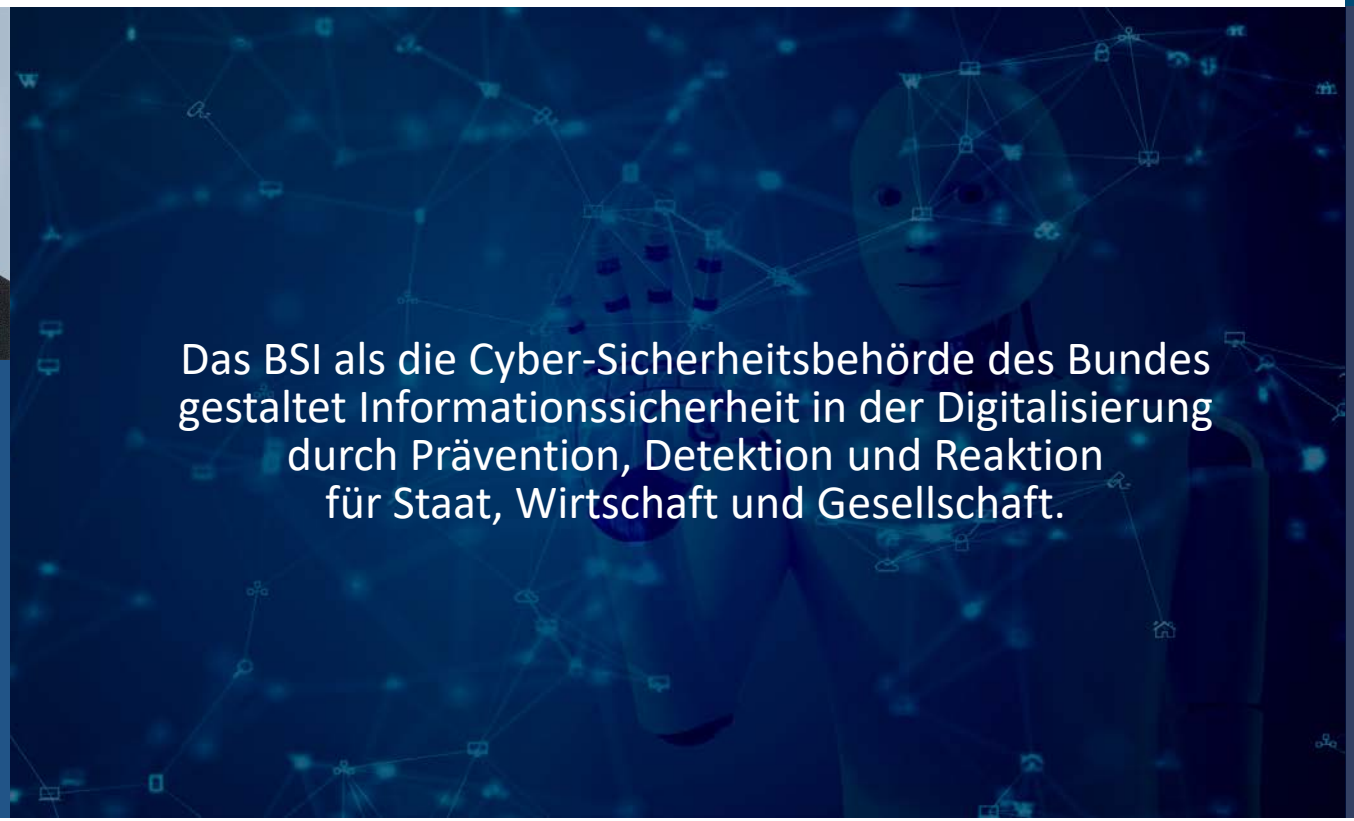
Contact:

Daniel Loevenich

daniel.loevenich@bsi.bund.de
Tel. +49 (0) 228 9582 5395
Fax +49 (0) 228 10 9582 5395
Mobil +49 (0)171 30 29 824

Bundesamt für Sicherheit in der Informationstechnik (BSI)
Referat TK 23: Grundsatz, Strategie und Nachweise in der Künstlichen Intelligenz

Godesberger Allee 185-189
53175 Bonn
www.bsi.bund.de



Bundesamt
für Sicherheit in der
Informationstechnik