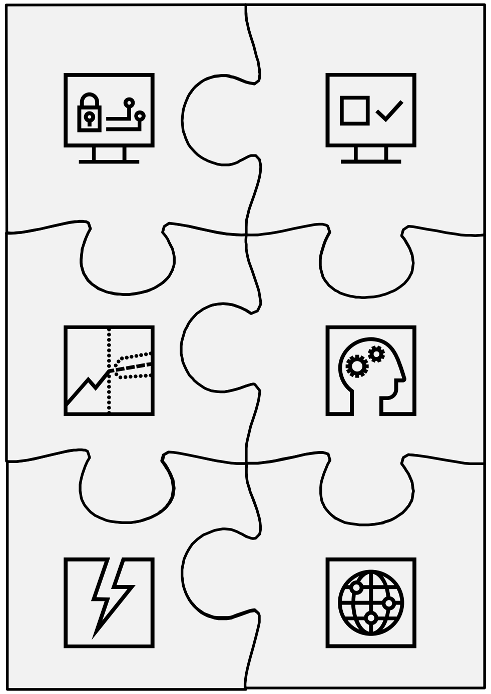
Conformity Assessment meets AI: Challenges and Concepts

PricewaterhouseCoopers Germany

October 2023

Conformity assessments are needed for establishing trust between relevant stakeholder of Artificial Intelligence

The **risk dimensions** of artificial intelligence require conformity…



Security & Robustness

Data Quality

Performance & Functionality

Explainability

Reliability

Bias & Fairness

… to **promote** trust and ensure safety when implementing and using AI.

User trust
Enables users to identify trustworthy AI products for purchase and usage.

Legal compliance and liability
Handle liability issues when operating AI products and services

Value chain certainty
Enables organizations to identify trustworthy parties in their value chain.

Investment certainty
Gives investors the ability to identify secure and trustworthy investment options

The current circumstances make it difficult for companies and auditors to carry out to demonstrate conformity

Increasing complexity of AI

Technological complexity

The sophistication of AI models through continuous learning and complex algorithms has increased the needed expertise.

Value chain complexity

AI application possibilities along value chains has come with challenges for tracking and auditing the way use cases cause risks.

Missing common frameworks

Missing evaluation methodology

The quick rise of AI applications has not been accompanied by the establishment and adoption of common regulatory frameworks.

Missing best practices

There are no best practices frameworks covering all relevant areas for artificial intelligence

A wide array of legal frameworks from governments and the industry aim to establish guidance for auditing AI

Excerpt

EU AI Act	BSI AIC4	NIST AI RMF	AI Verify	ISO/IEC	RAI Standard v2
Launched in 2024	**Published in 2021**	**Published in 2023**	**Published in 2022**	**Published in 2022**	**Published in 2022**
Mandatory rules for high risk AI applications within the EU	Criteria Catalog for auditing cloud-based AI products	Voluntary framework containing best practices for operating AI products	Voluntary AI governance framework for private sector	Guidance for the organizations to enable the safe and efficient use of AI	Concept for responsible AI
Classification of AI products into risk categories	Contains best practices and concrete audit procedures to safeguard AI products	Risk Management Framework for AI products	Testing framework and software toolkit	E.g. ISO/IEC 38507:2022 – Governance implications of the use of AI	Contains best practices and internal guidelines for operating AI responsibility

The EU AI Act focuses on a risk-based approach to AI, which is expected to come into force in the near future

Timeline

First draft
2021

Vote
2023

Into Force
2024

Application
2025

Risk categories

Unacceptable-risk AI
Prohibited AI systems (Article 5)

High-risk AI
Conformity assessment (Article 6)

Foundation models and GenAI
Specific requirements (Article 28b)

Extra transparency obligations
Obligations (Article 52)

AI with minimal or no risk
Allowed without restrictions

Type of conformity assessment

Internal Control

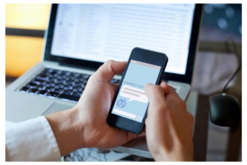
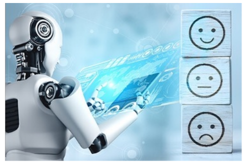
Self verification of quality management system, technical documentation and related processes

Notified bodies

Conformity assessment with the **involvement of an external notified body**

The fines in case of non compliance are up to 40M EUR or 7% of annual revenue.

AI use cases must comply with the requirements of the EU AI Act according to their respective risk category

	Without obligations	Transparency Obligations	AI system with high risk	Foundation models, GenAI	Unacceptable risk
Examples	 Spamfilter  Support chatbots	 Deepfakes  Emotion recognition	 Selecting job candidate Credit scoring	ChatGPT Dall-E	 Social Scoring Biometric Identification
Implications	Code of conduct	Notification about AI interaction	Conformity assessment	e.g. Data- and lifecycle management	Prohibited

Validation of AI systems need to be performed depending on the required risk assurance levels

Assurance Level

High

Risk Assurance Level	Implication	Link to the EU AI Act
Technical validation	Testing with the involvement of technical tools	The **conformity assessment** as described in the EU AI Act is required to be performed
Certification	Assessment with the involvement of an accredited party	with the involvement of a **notified body**
Attestation	Process and control testing by external party	or
Self Assessment	Validations by providers (or internal audit)	based on **internal control.**

Low

Case Study attestation: To enable conformity assessments, audit criteria can be used for designing a control framework

BSI AIC4 criteria

Excerpt

Security and Robustness

SR-05 – Based on the mitigation decisions for concrete threat models […], the AI model(s) are tested by implementing attacks to exploit identified vulnerabilities. The attacks tested are documented including […].

Performance & Functionality

PF-02 – The AI service provider assigns personnel to continuously compute and monitor the performance metric(s) defined in PF-01. […] reports on the performance of the service are communicated […].

Data Quality

DQ-03 – The quality of gathered data is continuously assessed […]. Corrective measures are in place to ensure stable data quality. The steps undertaken during data assessment are documented […].

Generic controls for compliance with the criteria

Illustrative

The AI **service** testing **team** conducts model robustness tests once a year. The test team uses different attack methods based on the attacker's objective, capability, and knowledge. The test processes and results under the preceding attacks are presented in the robustness test report. […]

The AI **operations team** is responsible for continuously monitoring the AI service against defined AI performance criteria in order to identify any deviation at the earliest possible stage and take appropriate countermeasures. Once a quarter, the AI Operations team reports to the responsible management […]

The AI **service operations manager** performs monthly assessments of data used for training and development of the AI service. The manager randomly selects samples from the annotated data to determine the data quality. Identified quality deviations are assessed and follow-up activities are initiated. [...]

Case Study attestation: Controls and the related evidence are required to demonstrate compliance

Excerpt

Criterion SR-05: Testing of Model Robustness

Attack resilience testing
Based on the mitigation decisions for concrete threat models for the AI model(s) within the scope of the AI service (e.g. based on adversarial attacks or privacy attacks) derived from the risk exposure assessment in SR-02 and SR-03, the AI model(s) are tested by implementing attacks to exploit identified vulnerabilities.

Attack documentation
Specifications of the implementation and configuration of the tested attacks are documented, including the results of the tests.

System response documentation
The attacks tested are documented including the observed system behavior of the AI service. Threat models, attack vectors and identified vulnerabilities are followed up as specified in SR-06.

Control to cover criterion SR-05

Yearly standardised testing
The AI service testing team conducts comprehensive model robustness tests once a year. The test team uses different attack methods based on the attacker's objective, capability, and knowledge. The test processes and results under the preceding attacks are presented in the robustness test report.

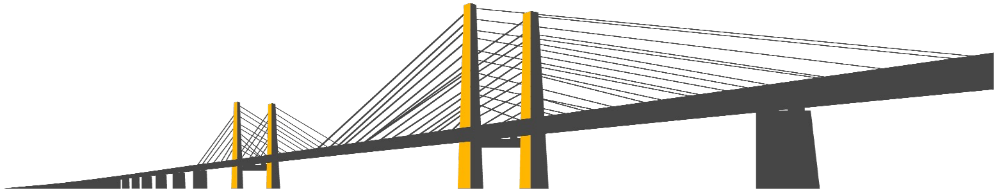
Analysis and vulnerability mitigation
The report summarizes and analyzes the results of each attack, provides an overall conclusion, evaluates whether the model has robustness risks, and defines measures to continuously track risks.

Evidence for the implementation of the control

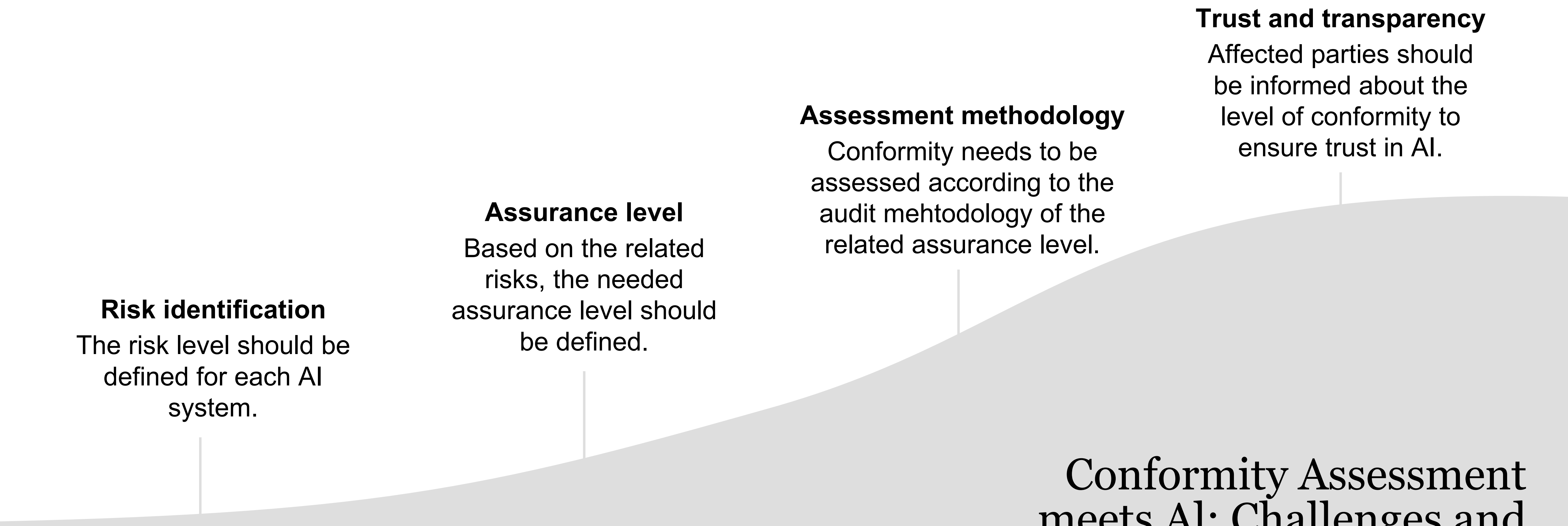
Overview of the development process
Evidence that robustness testing is an integral part of the AI development process.

Tool Screenshots
Code evidence showing that the model of the service in scope has been uploaded to the Robustness Tool and tests have been triggered.

Model Robustness Test Report
Result report of the robustness tests documenting which tests were carried out and what the results were.



For a conformity assessment, a risk-based approach according to the use-case specific properties is necessary.



Risk identification

The risk level should be defined for each AI system.

Assurance level

Based on the related risks, the needed assurance level should be defined.

Assessment methodology

Conformity needs to be assessed according to the audit mehtodology of the related assurance level.

Trust and transparency

Affected parties should be informed about the level of conformity to ensure trust in AI.

Conformity Assessment meets AI: Challenges and Concepts

Q&A

Thank you.

pwc.de

Hendrik Reese
Partner

+49 151 704 23 201

hendrik.reese@pwc.com