

Key to Ensuring Continuous Compliance: Assessing the Residual Risks of AI Systems/Products with Embedded Software

Challenges and Opportunities

Valentin Nikonov, International Expert on Risk Management,

Vice Chair UNECE WP.6 GRM

UNECE Disclaimer

- The author and the speaker of this presentation confirm that they have authorization to use all photos and visual elements.
- The material is either copyright-free or the author / speaker holds the necessary copyright.
- The UNECE will remove any material from its events and supporting websites if there is unlawful use of copyrighted material.
- The author / speaker takes responsibility for any infringements on copyright and holds the UNECE harmless to this effect.

Objective and contents

Key messages

To ensure continuous compliance of products with embedded software/AI systems, it is essential to assess the residual risk associated with each product

Assessing the residual risk of products with embedded software/AI systems is a challenging task

Addressing these challenges requires introducing new tools in Quality Infrastructure and International Cooperation

Contents

1. WP.6
Recommendation R
and Regulatory
Frameworks for AI
Systems

2. Compliant AI
Systems, Residual Risks
and WP.6
Recommendation S

3. Challenges in Evaluating the Residual Risk of
Products with Embedded Software (AI systems)

4. Existing Frameworks
for the Safety
Evaluation of AI
Systems

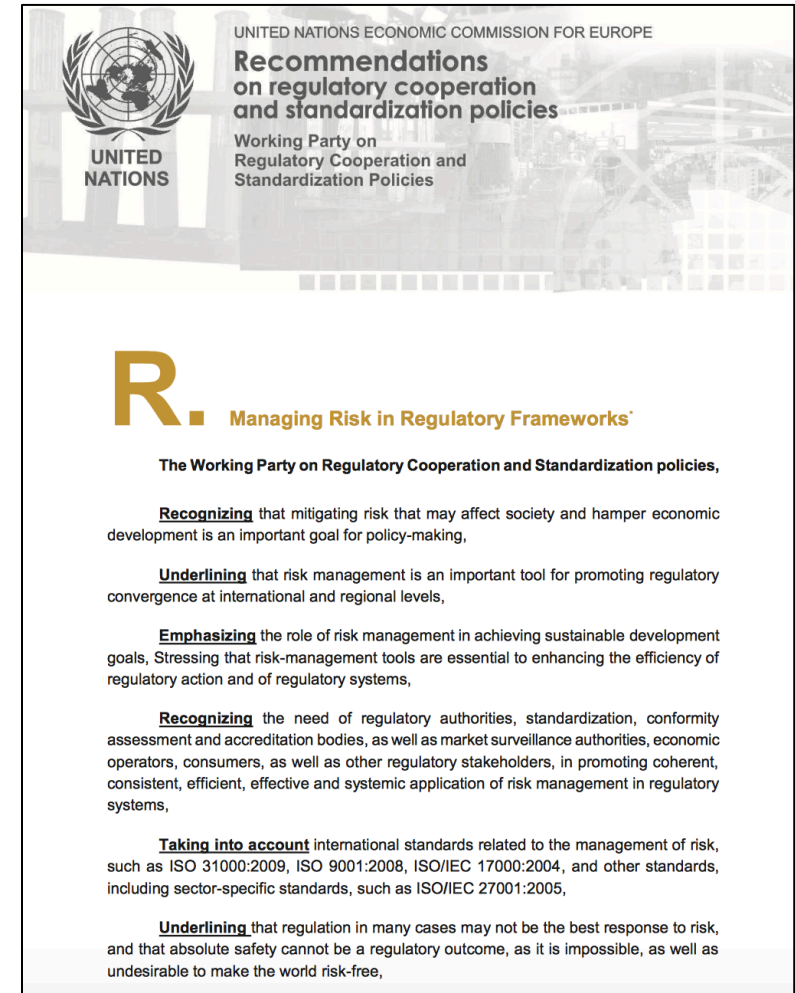
5. Conclusion and
Opportunities

WP.6 Recommendation R and Regulatory Frameworks for AI Systems

Challenges of Regulating Black-Boxes

Regulatory Frameworks for AI Systems are even more risk-based

- WP.6 Recommendation R (2011) describes a Risk-Based Regulatory Framework and presents regulation as a risk mitigation tool
- It recommends, among other things, that:
 - “**All functions** of the risk management process should be consistently described in legislation that lays out the regulatory framework at a general level or for a specific sector”
 - “Regulatory authorities should establish, implement and maintain, a process for determining, analyzing, reviewing and monitoring an **acceptable level** of risk within a regulatory framework”
- AI systems are black (or grey) boxes
- Regulatory Frameworks for AI Systems are **even more** risk-based than those for “traditional” products



Regulating Traditional Markets vs. Regulating AI Systems: difference in approaches

Traditional, Deterministic Products/Systems

- A Regulator can establish requirements for:
 - **Products characteristics**
 - Related processes
 - Production methods
- Product characteristics refer to attributes of a product (such as width, weight, etc.)
- Regulation describes the regulated product itself

Regulating AI Systems – black/grey boxes

- Regulations establish requirements for AI system provider/other stakeholders **to mitigate risks** of a system
- Regulations require **the residual risk of an AI system to be acceptable**
- Regulation sets out risk management processes and mitigation methods, applied to a system

Example of a Regulatory Framework: EU AI Act

1. Regulation sets out requirements for a risk management process:

2. The risk management system shall consist of a continuous iterative process run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic updating. It shall comprise the following steps:
 - (a) identification and analysis of the known and foreseeable risks associated with each high-risk AI system;
 - (b) estimation and evaluation of the risks that may emerge when the high-risk AI system is used in accordance with its intended purpose and under conditions of reasonably foreseeable misuse;
 - (c) evaluation of other possibly arising risks based on the analysis of data gathered from the post-market monitoring system referred to in Article 61;
 - (d) adoption of suitable risk management measures in accordance with the provisions of the following paragraphs.

2. Regulation describes risk mitigation measures for developing AI systems, such as:

- Data and data governance,
- Technical documentation,
- Record keeping,
- Quality management system, etc.

3. Regulation establishes requirements for acceptability of the **residual** risk:

any residual risk associated with each hazard as well as the overall residual risk of the high-risk AI systems is judged acceptable, provided that the high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse. Those residual risks shall be communicated to the user.

Compliant AI Systems, Residual Risks and WP.6 Recommendations

AI System is compliant when its level of residual risk is tolerable

Non-compliance risk of an AI System

- WP.6 Recommendation S generalizes the concept of non-compliance risk
- According to Recommendation S, the evaluation of the non-compliance risk should take into account:
 - **Consequences** of non-compliance (of an AI system)
 - **Probability** of non-compliance (of an AI system)
- Consequences of non-compliance can be determined for groups of products/systems (example of high risk systems)
- Probability of non-compliance will differ from system to system, from product to product

High risk

AI systems identified as high-risk include AI technology used in:


- critical infrastructures (e.g. transport), that could put the life and health of citizens at risk;
- educational or vocational training, that may determine the access to education and professional course of someone's life (e.g. scoring of exams);
- safety components of products (e.g. AI application in robot-assisted surgery);
- employment, management of workers and access to self-employment (e.g. CV-sorting software for recruitment procedures);
- essential private and public services (e.g. credit scoring denying citizens opportunity to obtain a loan);
- law enforcement that may interfere with people's fundamental rights (e.g. evaluation of the reliability of evidence);
- migration, asylum and border control management (e.g. verification of authenticity of travel documents);
- administration of justice and democratic processes (e.g. applying the law to a concrete set of facts).

Example of the EU AI Act: High Risk systems
(based on consequences of non-compliance)

Risk classification of AI systems and residual risk

- The risk classification of AI systems (as referred to in legislation) can be based solely on the consequences of product/system non-compliance
 - Using a non-compliant AI system in critical infrastructures will lead to more severe consequences than using of a non-compliant chatbot
- Probability of non-compliance, representing the likelihood that an AI system will cause harm, is the residual risk of an AI system
- The approaches described in Recommendation S can be adopted for evaluation of residual risk

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE



**Recommendations
on Regulatory Cooperation
and Standardization Policies**
Revised Edition

**S Applying predictive risk management tools
for targeted market surveillance¹**

The Working Party on Regulatory Cooperation and Standardization Policies,
Emphasizing that achieving absolute safety cannot be the goal of a regulatory system,
Noting that excessively stringent controls can create unnecessary barriers to trade,
Recognizing the importance of ensuring that products on the market (including imported goods), physical infrastructure, commercial and industrial facilities are compliant and safe so as to protect consumers, citizens and the environment
Emphasizing the importance of applying predictive risk assessment tools for planning the activities of market surveillance/compliance authorities at the “before an accident”/“before the non-compliance reported” stage,
Stressing that risk-based surveillance frameworks should help avoiding:

- Excessive controls on low risk products and
- Omitted or insufficient controls on high risk products

Recognizing that authorities need to efficiently allocate limited resources and that risk-based targeted surveillance on products on the market (as well as on installations and facilities) provide an important means to that end,
Aiming to provide guidance in the use of predictive risk management techniques so as to increase the efficiency of the existing risk assessment tools and data sharing platforms,
Aiming to complement the existing risk assessment tools applied by market surveillance authorities,
Recommends that: Authorities plan surveillance activities on the basis of the evaluation of the non-compliance risk of products/businesses within their jurisdiction. The evaluation of the non-compliance risk should reflect:

- How dangerous a certain product/business entity is when it is non-compliant to standards,
- What is the probability that a non-compliant product of this type is present on the market.

What is a compliant AI system?

- AI system is a black/grey box: functionality is unknown/partly unknown
- It is impossible to “look inside” to check how it works
- System is stochastic, not deterministic



- A compliant AI system is developed/operated under conditions that mitigate/minimize/eliminate/reduce risks
 - Relatively easy to inspect during conformity assessment
- A compliant AI system maintains an acceptable/tolerable level of residual risk
 - Evaluating the residual risk poses a significant challenge

Conformity Assessment plays a crucial role within a Risk-Based Regulatory Framework

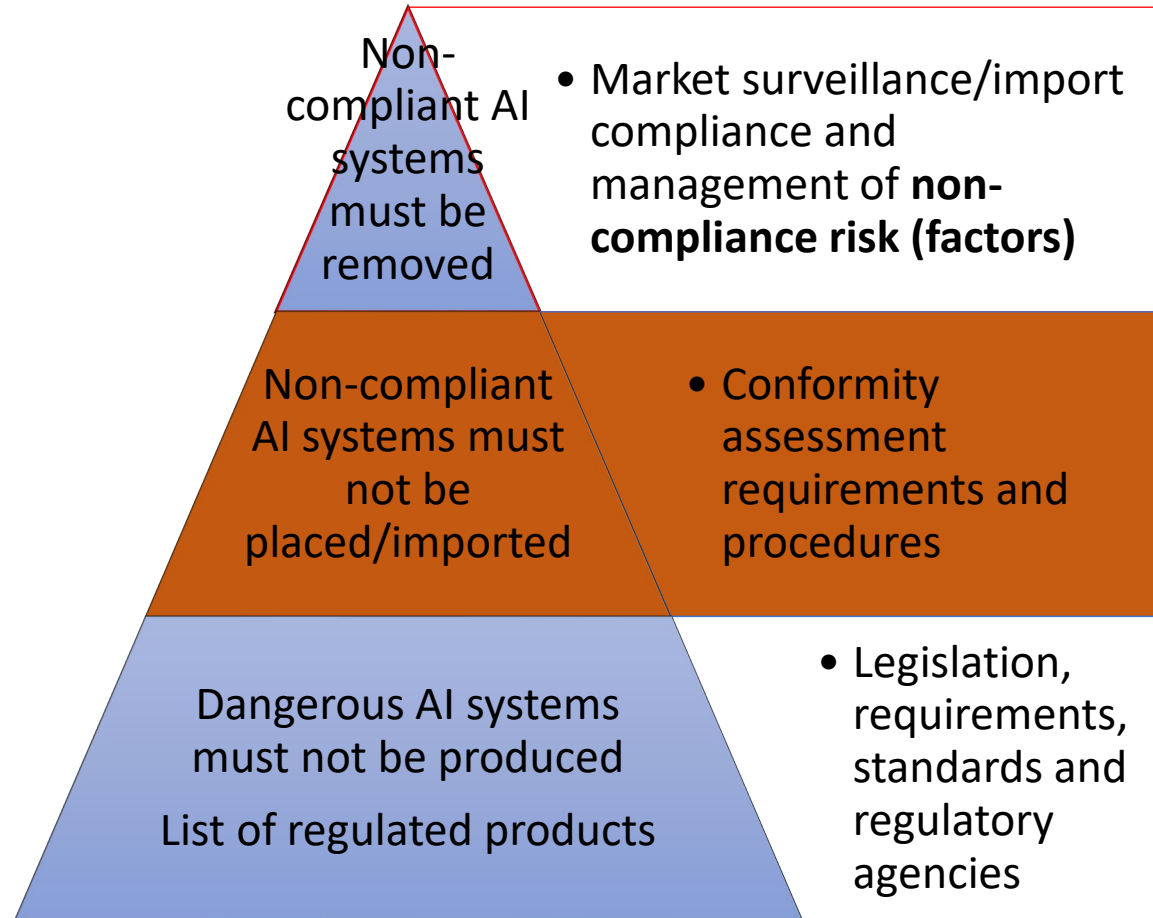
On which products to focus in surveillance/import compliance?



Which CA schemes are adequate for different products/systems?



Are regulatory requirements proportionate to risks they were set out to address?



Setting regulatory objectives: no absolute safety

Management of assets

Risk identification

Risk evaluation

Choosing risk treatment strategies

Contingency planning

How much risk is acceptable/tolerable?

Conformity Assessment of AI systems: evaluating the residual risk

- Conformity assessment aims to prevent products with unacceptable level of non-compliance risk from being placed on the market
- Different regulatory authorities may have different criteria for determining what constitutes an unacceptable level of non-compliance risk
- Criteria include various combinations of consequences and probabilities of non-compliance
- The easiest case of unacceptable risk:
 - A high risk system (high consequences of non-compliance)
 - High probability of non-compliance (residual risk)
- **Conformity assessment of AI systems requires evaluation of residual risk of each AI system**

Consequence rating ↑	a	III	III	II	I	I
	b	IV	III	III	II	I
	c	V	IV	III	II	I
	d	V	V	IV	III	II
	e	V	V	IV	III	II
		1	2	3	4	5
		Likelihood rating →				

Evaluating the residual risk: two main questions

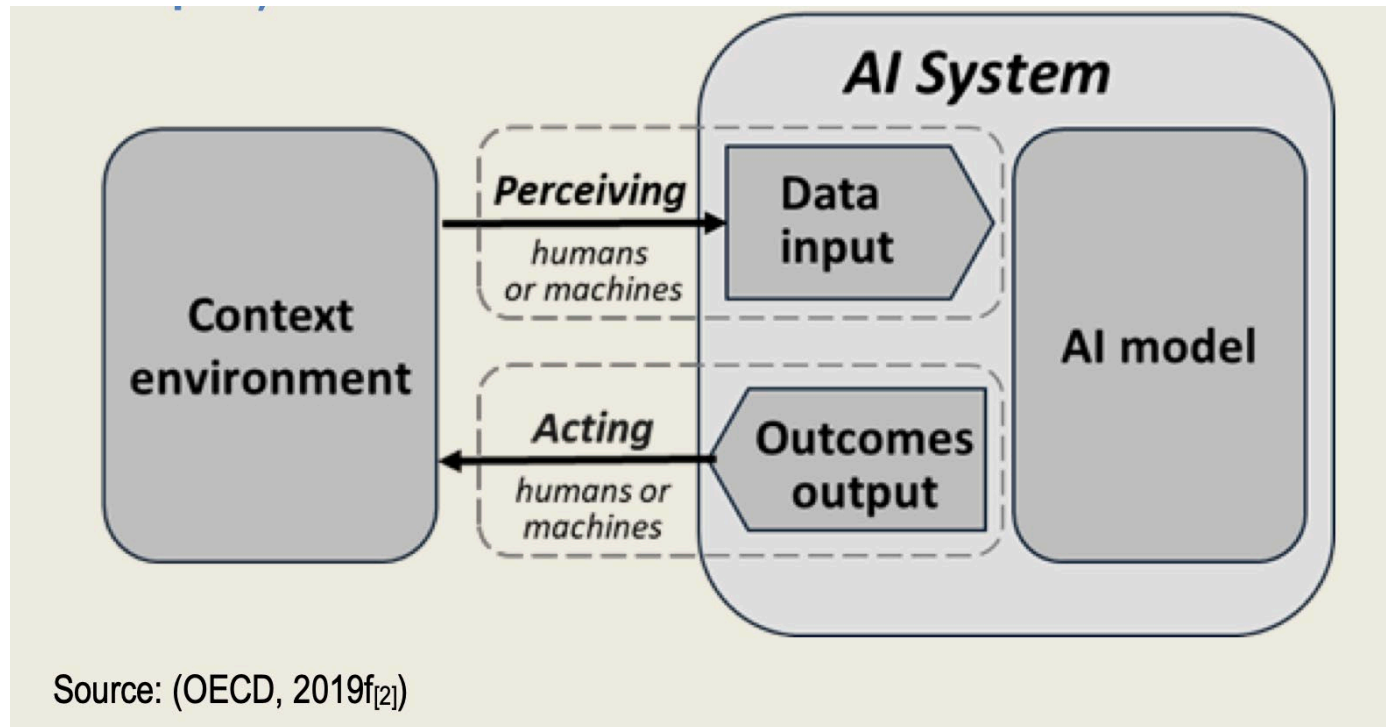
In which situations/scenarios, an AI system will fail/cause harm?

How likely are these scenarios?

Challenges in Evaluating the Residual Risk of Products with Embedded Software (AI systems)

Challenges of Evaluating Residual Risk: we can see results only

Conceptual view of an AI System (OECD):

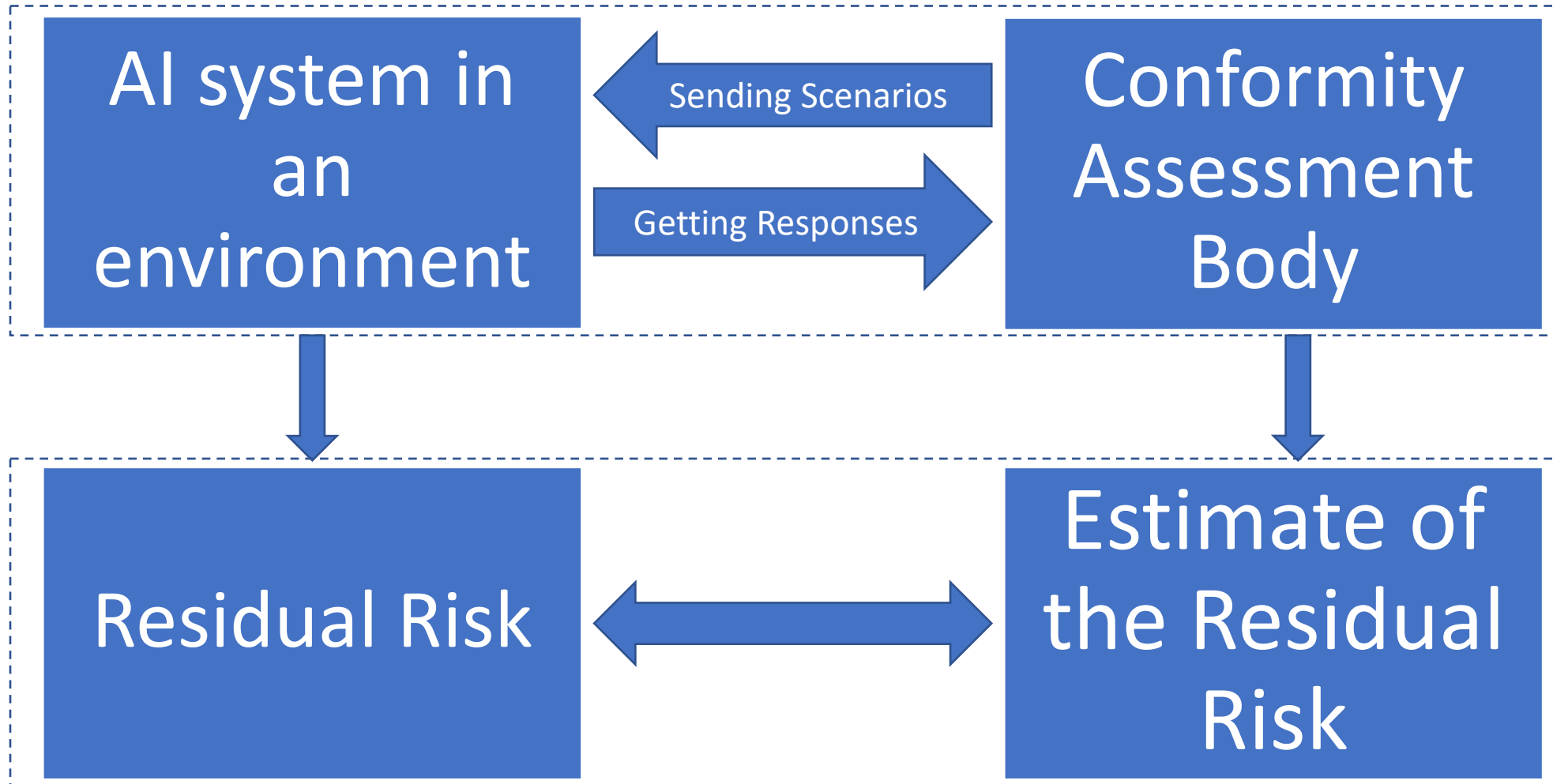


1. Conformity Assessment Body knows the data (scenarios) a system inputs

2. Conformity Assessment Body knows the data a system outputs (its behavior given certain scenario)

3. Conformity Assessment Body doesn't know the AI model itself well enough to check it and to be sure it is safe

A framework for assessing conformity (evaluating residual risk) of an AI system



Regulatory Approval of an AI system: regulatory and conformity assessment challenges

Regulatory challenge:
How safe is safe enough?
What is the acceptable level of the residual risk?

- Challenging task for any Regulatory Authority
- One of the approaches – GAME principle, Globalement au moins equivalent

Tolerable level of residual risk

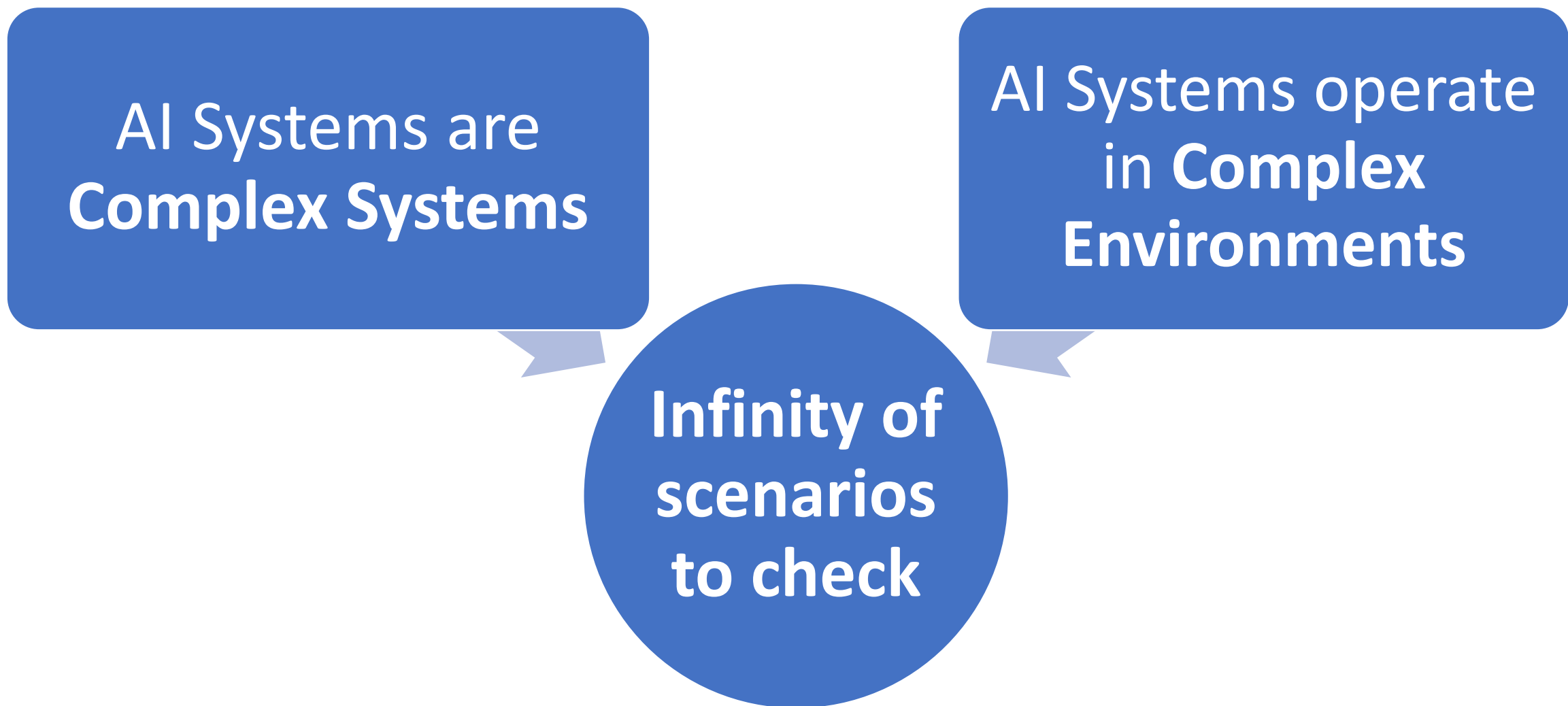
Conformity Assessment Challenge:

How much testing will be enough to prove that a product is safe?

- Challenging task for any Conformity Assessment Body
- One of the approaches – scenario-based simulation to ensure regulatory compliance

Conformity Assessment Challenge: so many scenarios to check

AI Systems are
Complex Systems

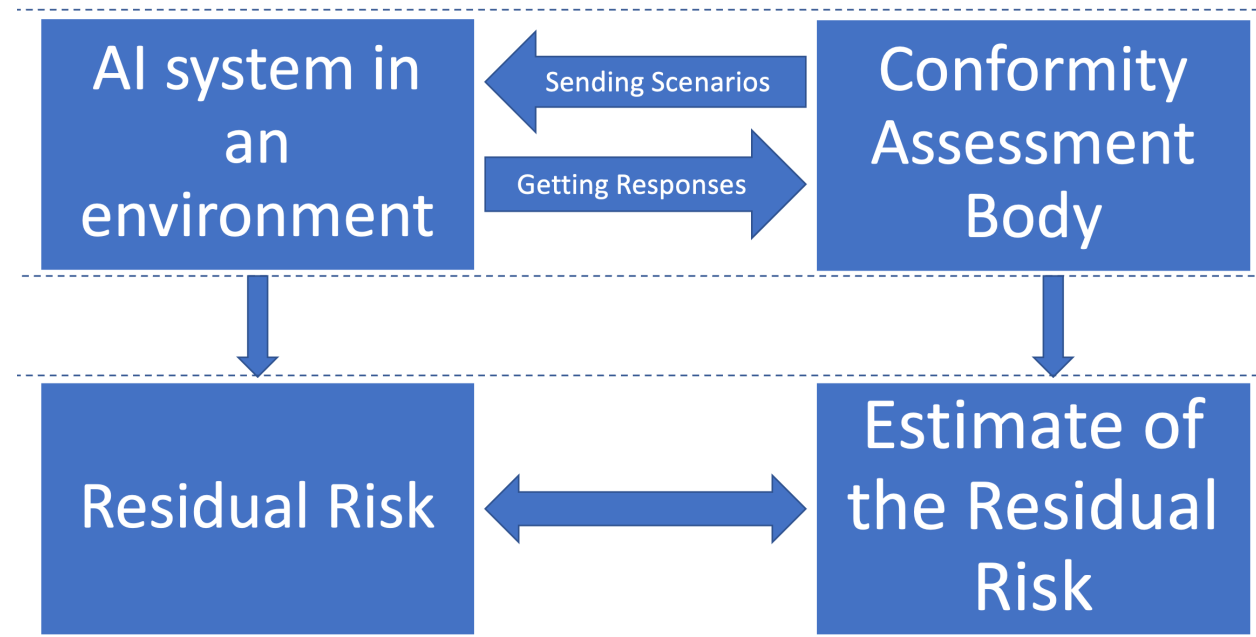


AI Systems operate
in **Complex
Environments**

**Infinity of
scenarios
to check**

Critical considerations/key questions in Conformity Assessment of AI systems

- **How to test a product:**
 - Physical test or simulation in a lab?
- **How to choose which scenarios to test:**
 - Which scenarios are most likely to happen in reality?
 - Which scenarios are most dangerous?
- **How to evaluate the responses of the tested product:**
 - How to “translate” the behavior of an AI system in metrics?
- **Can we trust the results:**
 - How can we know that we tested enough?
 - Can we trust our estimates of the residual risk?

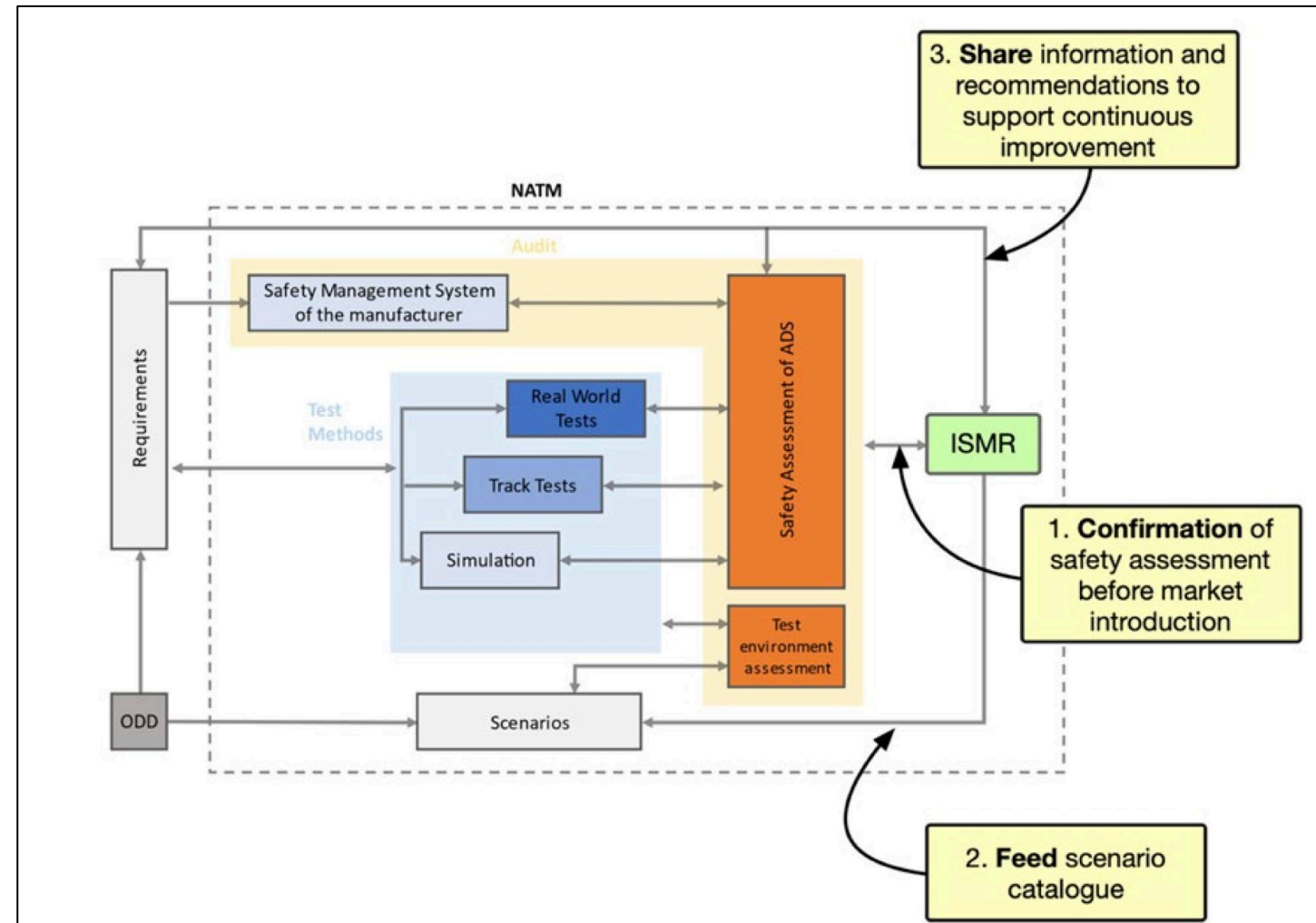


Existing Frameworks for the Safety Evaluation of AI Systems

One example

Scenario-based approaches for evaluation of residual risk for (most) complicated AI Systems

- Informal Working Group on Validation Methods for Automated Driving, UNECE WP.29
- New Assessment/Test Method for Automated Driving Guidelines for Validating Automated Driving Systems
- Similar approaches could be developed for products/systems within the scope of WP.6



Conclusion and Opportunities

1. Ensuring continuous compliance of AI systems/products with embedded software requires the evaluation of their residual risk
2. This evaluation is an indispensable part of the conformity assessment process
3. Evaluation of residual risk is a challenging task, which is key to ensuring safety
4. No matter if performed by a system developer or by a third-party, addressing the challenges necessitates application of scenario-based approaches and simulation methods within Quality Infrastructure
5. These methods are already being developed in several industries facing the challenges of regulatory approval of black/grey boxes

Conclusion and Opportunities

6. Regulatory cooperation in the approval of AI systems is essential for ensuring safety and facilitating trade
7. The focus of regulatory cooperation can encompass:
 1. Development of scenario databases for different product groups to be used in conformity assessment
 2. Establishment of common approaches for determining the acceptable level of risk
 3. Development of methodologies for performing conformity assessment procedures and evaluating the residual risk of products
8. WP.6 can serve as a platform for regulatory cooperation in building conformity assessment frameworks for product with embedded software within its scope
9. WP.6 GRM can be a platform for the development of methodologies for evaluation of residual risk

Key to Ensuring Continuous Compliance: Assessing the Residual Risks of AI Systems/Products with Embedded Software

Challenges and Opportunities

Valentin Nikonov, International Expert on Risk Management,

Vice Chair UNECE WP.6 GRM