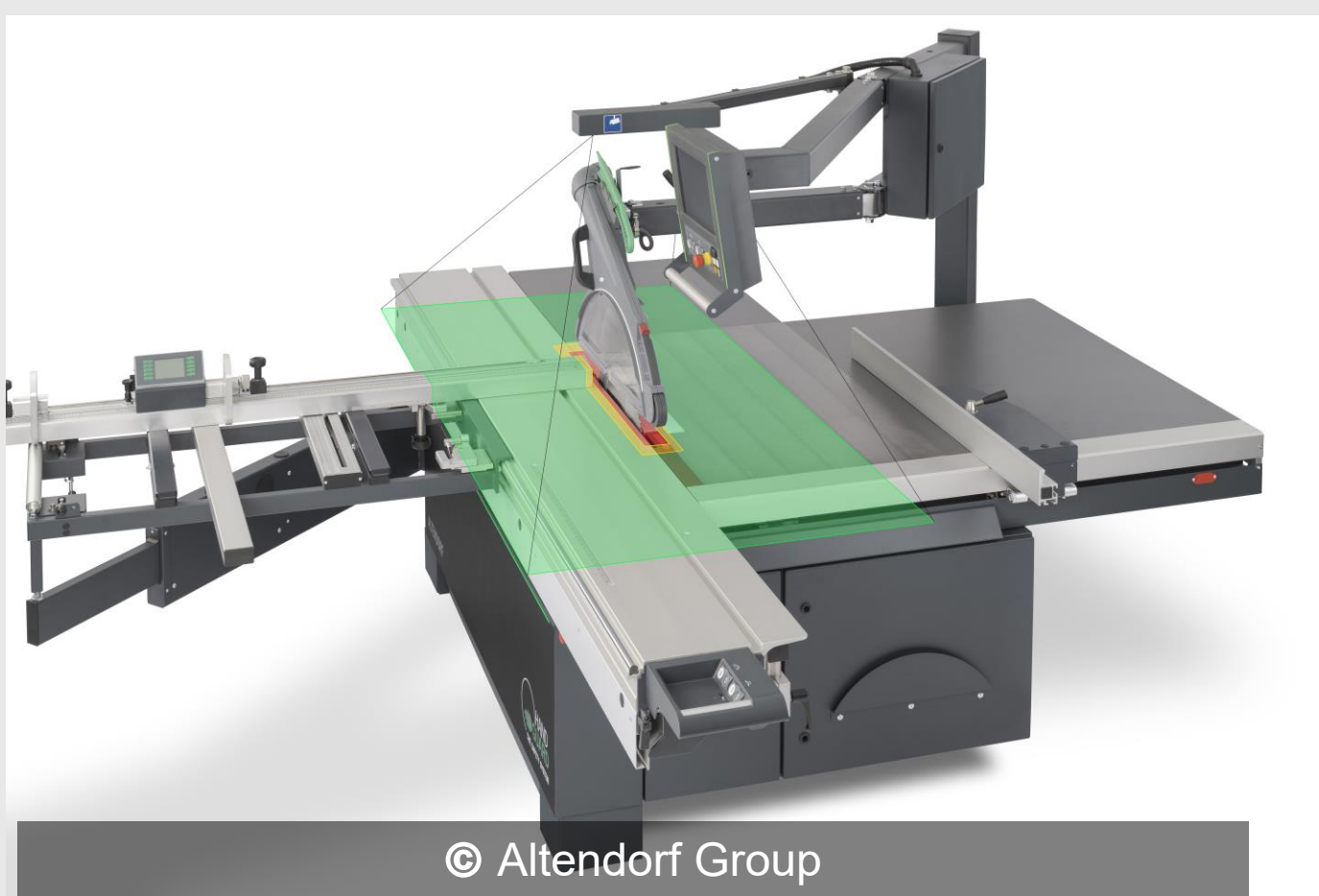


Towards Safe Integration of AI in Machinery: Challenges and Evaluation Approaches

When do we need to consider *functional safety* in the context of AI?

- **AI in safety-relevant development:** When AI is used in the development of safety-critical components (e.g., software for machine control systems).
- **AI in safety-critical control:** When AI directly assumes responsibility for safety-critical control functions.
- **AI as safety or assistance function:** When AI is deployed as a safety function or assistance feature within a machine.



AI can support many of these use cases, but it must never introduce new hazards!

What are the challenges?

AI-Systems are usually „black boxes“

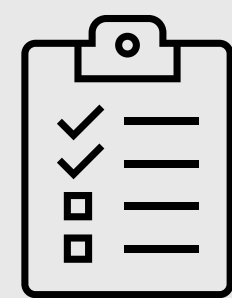
- The output of an AI system cannot be predicted for every possible input.
- It is not possible to fully reconstruct what the system has learned in detail.
- Extensive testing is required to ensure that the system operates robustly under as many conditions as possible.

Developers must extract and interpret requirements from applicable legislation (e.g., the EU AI Act).

There are no standardized testing or assessment procedures.

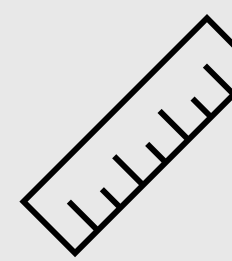
No specific requirements from standards and legislation yet

How does the IFA contribute to ensuring the safe use of AI in the machinery domain?

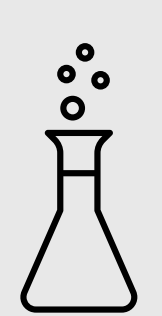


Established testing institute: The IFA is an established testing institute, particularly for emerging technologies, with extensive industry connections in the field of safety engineering

- Initial evaluations of AI-based assistance systems have been completed or are currently underway.
- Numerous advisory and informational discussions are conducted with manufacturers regarding the use of AI in machinery and the associated regulatory and technical requirements.
- The IFA leads a DGUV Test working group dedicated to developing testing principles for AI-based assistance systems.



The IFA actively participates in ongoing **standardization projects** related to AI and functional safety.



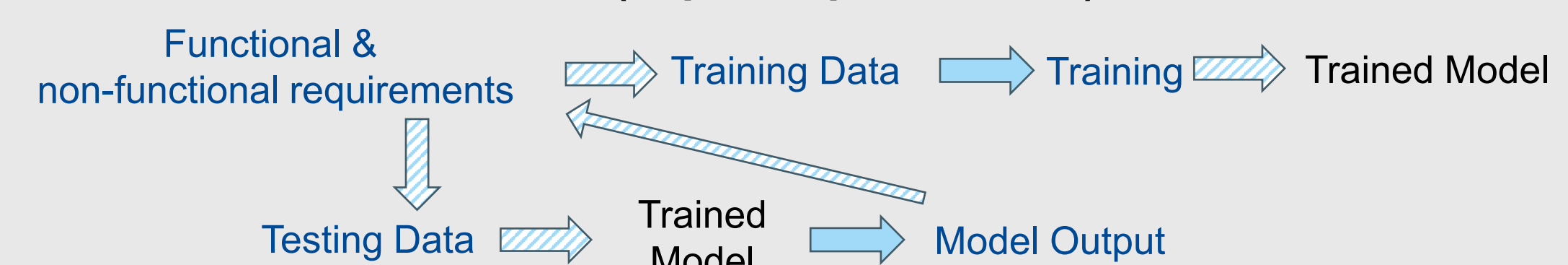
The IFA is **connected within the scientific community** on AI safety and **seeks collaborations with relevant stakeholders** to ensure that the testing and evaluation of AI systems is supported by rigorous scientific methodology.

What distinguishes the development and evaluation of AI from that of traditional software?

Classical Software (explicit specification)



Data-driven models (implicit specification)



- In traditional software, the intended behavior is explicitly specified and implemented through deterministic program code (e.g., rule-based if-then logic).
- AI models learn behavior from training data using statistical methods, meaning the specification is implicit and determined by data selection.
 - Which relationships the model ultimately extracts from the data is generally not fully traceable.
 - Evaluation is only possible through extensive test datasets that cover a broad range of operating conditions

What do we need to test AI?

To assess the quality of an AI system, we can examine many different stages of its development — such as the training process, data quality, or specialized methods used to improve the system. But to truly evaluate the robustness of an AI system, we need

Data, Data, Data!

- When we want to test an AI system with highly specific inputs, **synthetic data or data augmentation** can be useful to isolate and analyze potential influencing factors.
- **However:** To ensure that an AI system truly performs as expected within its defined Operational Design Domain (ODD), we additionally require a large volume of diverse real-world test data that **have not been used in the development process.**

It is essential for testing facilities to start building (and maybe collaborating on) their own, extensive test data sets, to ensure we have reliable, functional safe AI systems in the coming future!

