Fraunhofer IAIS

Dr. Maximilian Poretschkin | June 21, 2023

Trustworthy AI - Assessing the quality of AI-systems

Intelligent Systems that Work!

AI, Machine Learning and Big Data

Research in the paradigm of "hybrid AI“ / “triangular AI”

Partnership with the Excellence University of Bonn, part of *The Lamarr Institute for Machine Learning and Artificial Intelligence*

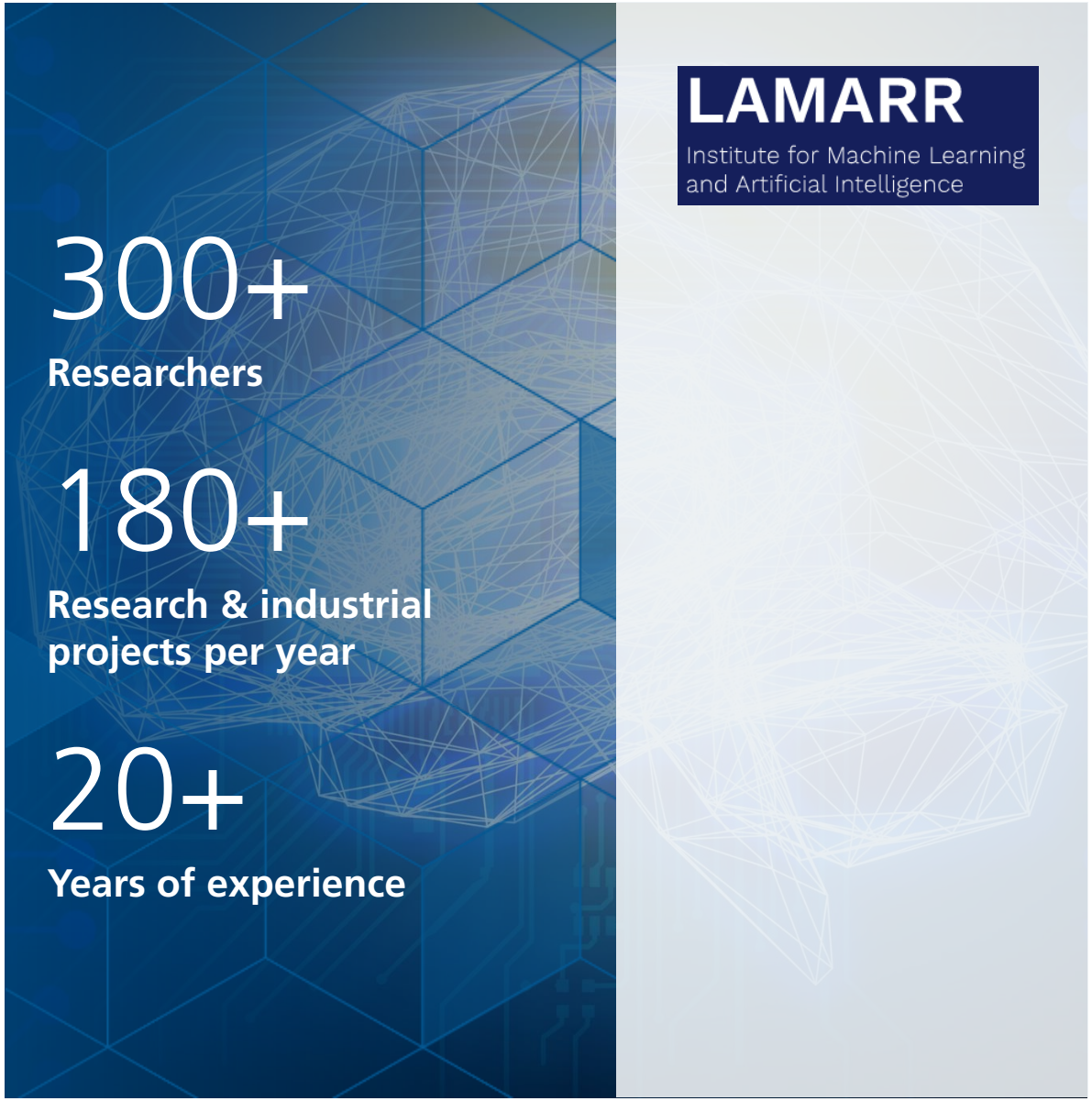
Comprehensive, immediately deployable, proven high-performance technology and IP portfolio

Consulting, 24-7 implementation, software, licensing, innovation partnerships, training

Customers and partners from DAX30 to medium-sized businesses

Network management KI.NRW, Fraunhofer Big Data and Artificial Intelligence Alliance, AI4EU

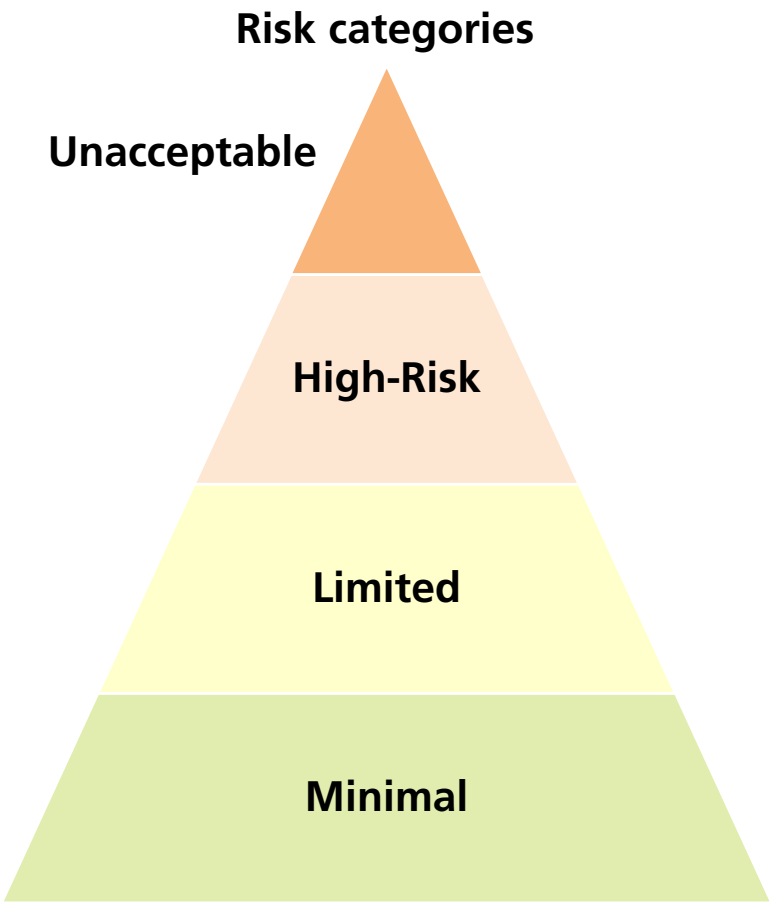
Particular focus on safeguarding and certification



Fraunhofer IAIS

Upcoming European AI regulation calls for AI conformity assessments

Regulation follows a risk-based approach, four risk categories defined



Description of AI system	Regulation	Example
Contravening Union values (e.g., fundamental rights)	Prohibited	AI-based social scoring …
High risk to the health and safety or fundamental rights of natural persons	Conformity assessment with specific rules required	Candidate selection…
Interaction with humans, detection of emotion or association based on biometric data, deep fakes	Transparency obligations	Chat bots …
Represent only minimal or no risk for citizens or safety	No regulation	Spam filter …

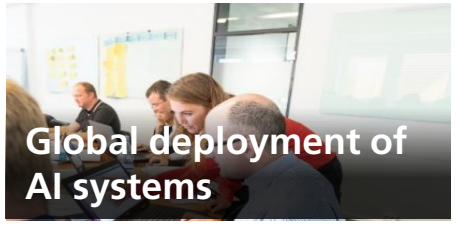
Motivation for AI assessments besides regulation

AI assessments can provide strategic market advantages

Building internal trust

Are the AI-based recommendations comprehensible and trustworthy?

Can malfunctions cause significant (personal and/or financial) damage?



Is the AI system reliable enough to be rolled out globally?

Building external trust

Can a competitive advantage be generated through proven technical reliability?

How can a trusted brand be maintained for products with AI components?

Understanding risks

Does the purchased AI solution meet the required characteristics?

Does a company takeover entail technical risks? Does an acquired AI solution meet the expected requirements?

Are AI risks recorded and assessed transparently? Are internal AI guidelines implemented?

Risk transfer

Can proof of technical reliability reduce the insurance premium?

Can the residual risk be covered by AI insurance?

Image sources: fizkes - stock.adobe.com, Nataliya Hora - stock.adobe.com, Fraunhofer IAIS, tanaonte - stock.adobe.com, Jacob Lund - stock.adobe.com, Looker_Studio - stock.adobe.com, amnaj - stock.adobe.com, Song_about_summer - stock.adobe.com, Zerbor - stock.adobe.com, TimeStopper - stock.adobe.com

Fraunhofer IAIS

Fraunhofer AI Assessment Catalog

Guidelines for a structured evaluation of AI to develop trustworthy AI

Step 1: Risk analysis

- Comprehensive risk analysis along the dimensions of fairness, autonomy and control, transparency, reliability, safety and security and data protection

Step 2: Definition of targets

- Definition of objectives and - preferably measurable - target criteria to mitigate the risks identified in step 1

Step 3: Documentation of measures

- Guidance to systematically list measures along the lifecycle of the AI application to achieve the targets set in step 2

Step 4: Assurance argumentation

- Guidance to develop a stringent argumentation based on the measures of step 3 to demonstrate that the objectives formulated in step 2 have been achieved

Assessment Catalog is freely available at:
https://www.iais.fraunhofer.de/en/research/artificial-intelligence/ai-assessment-catalog.html

Areas of Application

The Assessment Catalog supports

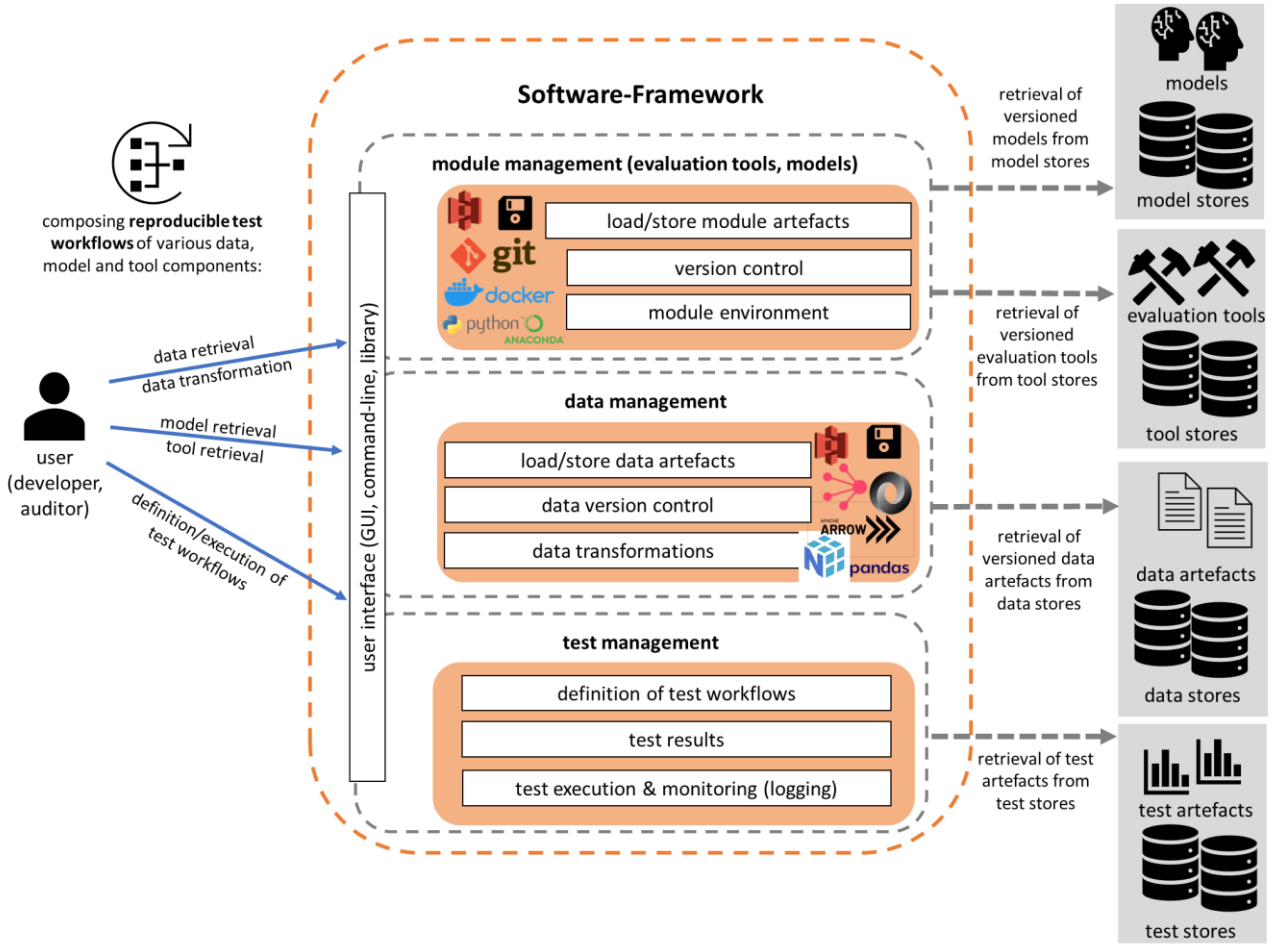
- Developers in the design
  and
- AI assessors in the evaluation and quality assurance

of AI applications.

ZERTIFIZIERTE KI
Qualität sichern. Fortschritt gestalten.

Automation of AI-assessments

Testing-framework integrates different AI-assessment tools



ZERTIFIZIERTE KI
Qualität sichern. Fortschritt gestalten.

Software-framework for reproducible and comparable tests

- Versioned storage of
  - Data
  - AI assessment tools (Toolsuite)
  - Models
  - Pipelines
  - Tests
- Interoperable, containerized module for models and AI assessment tools
- Cloud-compatible software-stack

Fraunhofer IAIS develops assessment service for Munich Re

Assessment service CertAI aims to increase acceptance of AI

Munich Re is building a new business area for quality assessment of AI solutions under the brand of »CertAI« – Fraunhofer IAIS is the technology partner

- Subject of the assessment are fully developed or already actively deployed AI systems
- Two assessment dimensions:
  Assessment of the process and assessment of the product. The results are a quality seal and a detailed assessment report
- Assessment service based upon the Fraunhofer IAIS »AI Assessment Catalog«
- Fraunhofer IAIS assists Munich Re with technical product assessments

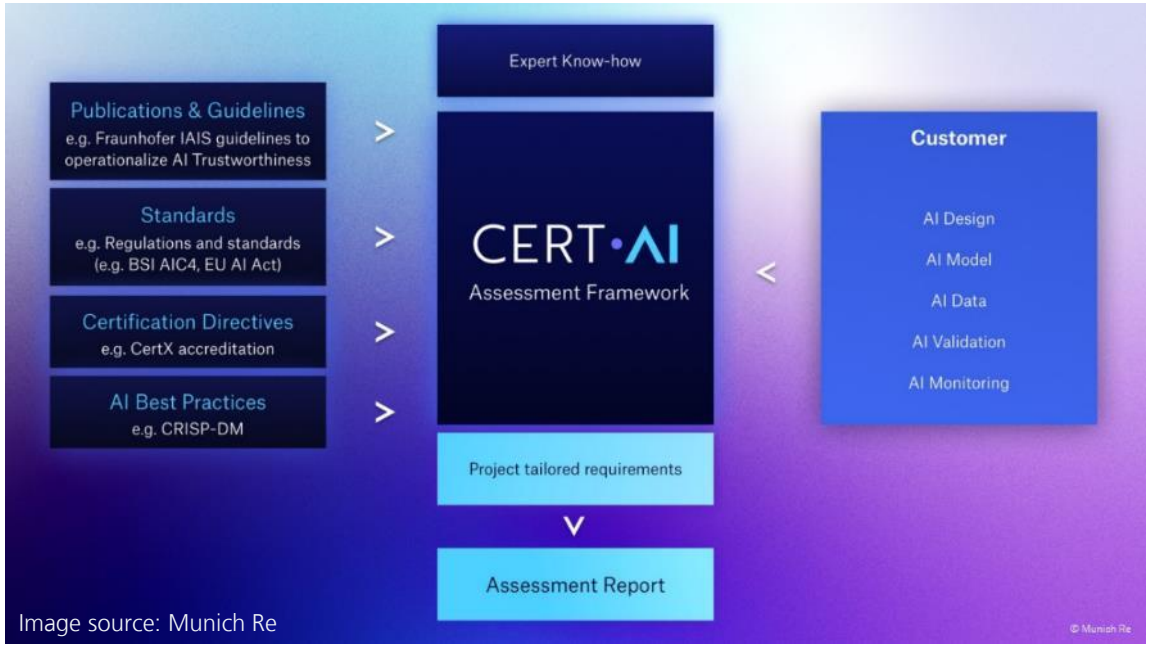


Image source: Munich Re

Fraunhofer IAIS Assessment Catalog sets standards for AI product assessments on the market

Fraunhofer IAIS

Many parallels with Japanese approach

More international cooperation needed

Machine Learning Quality Management Guideline

3rd Edition

January 20, 2023
(Japanese: August 2, 2022)

Technical Report DigiARC-TR-2023-01
Digital Architecture Research Center

Technical Report CPSEC-TR-2023001
Cyber Physical Security Research Center

Technical Report
Artificial Intelligence Research Center

National Institute of Advanced Industrial Science and Technology (AIST)

AIST

Fraunhofer IAIS