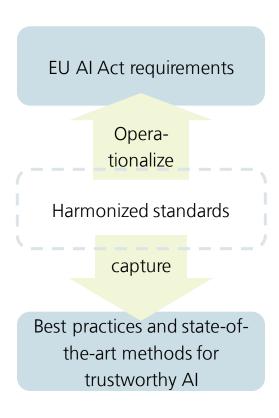
Recent works from the project ZERTIFIZIERTE KI

How to implement and demonstrate conformity with fairness-related AI Act requirements in practice?



Al fairness requirements in the EU Al Act and their technical evaluation

Joint work by ZERTIFIZIERTE KI project partners Universität Köln & Fraunhofer IAIS



Protection from discrimination is among the recitals of the EU AI Act

- How do the Al Act requirements address this objective?
- Is it sufficiently ensured?

See: Feldkamp, J., Kappler, Q., Poretschkin, M., Schmitz, A., Weiss, E. (2024). Rechtliche Fairnessanforderungen an KI-Systeme und ihre technische Evaluation – Eine Analyse anhand ausgewählter Kreditscoring-Systeme unter besonderer Berücksichtigung der zukünftigen europäischen KI-Verordnung. Heft 1, ZfDR 2024.



Fairness-related requirements in the EU AI Act are focused on the data

From abstract requirements to their technical implementation

EU Al Act requirements Operationalize Harmonized standards capture

Specific fairness-related technical requirements in the Al Act (Art. 10)

- **Data quality** (relevance, completeness, representativity, free of errors)
- Examination of data in view of possible biases, mitigation

Best practices and state-ofthe-art methods for trustworthy Al

Our approach: Review of computer science literature on how the above concepts/ requirements can be interpreted, measured and implemented in practice

See: Feldkamp, J., Kappler, Q., Poretschkin, M., Schmitz, A., Weiss, E. (2024). Rechtliche Fairnessanforderungen an KI-Systeme und ihre technische Evaluation – Eine Analyse anhand ausgewählter Kreditscoring-Systeme unter besonderer Berücksichtigung der zukünftigen europäischen KI-Verordnung. Heft 1, ZfDR 2024.



Fairness-related requirements in the EU AI Act

Main findings

EU Al Act requirements

Operationalize

Harmonized standards

capture

Best practices and state-ofthe-art methods for trustworthy Al

Requirements in the Al Act need further specification



Various target concepts & implementation methods exist, with **different effects**

- on potential biases of data and models (e.g., representativity concepts)
- on their quality evaluation

Focus on data quality in the Al Act does not sufficiently protect from biases



No clear evidence of the general effectiveness of data quality measures for bias mitigation in ML models; SOTA approaches include measuring & mitigating biases in the ML model itself

Guidance is missing on how to take account of interaction between requirements



Potential trade-offs when implementing different fairnessrelated requirements (e.g., bias mitigation and correctness of data) and with other requirements (e.g., model accuracy)





Symposium on Scaling AI Assessments

Tools, Ecosystems and Business Models | Cologne 2024, September 30 - October 1

Reliable methodologies for Al assessments are acutely needed, but many central questions are still open.

The symposium connects **TIC-industry** professionals, startups, legal experts and research scientists to advance marketable and scalable solutions for Al assessment.

Two perspectives towards scalability

Operationalization of market-ready Al assessment



Risk and Vulnerability Evaluation

Business Models Based on Al Assessments

J. Implementing Regulatory Requirements

Testing tools & implementation methods for trustworthy Al

Infrastructure and Automation



Safeguarding and Assessment Methods



Systematic Testing



Contribute to the future of market-ready Al assessments and audits for trustworthy Al with practical expertise or scientific know-how and pre-register now for the Symposium in Cologne!



