How standardization brings AI trustworthiness into practice

Dr Sebastian Hallensleben

Head of Digitalisation & AI at VDE e.V.

Chair of CEN-CENELEC JTC 21

Co-Chair of OECD ONE.AI WG Risk Assessment

UNESCO Expert Group on AI Ethics

VDE and AI trustworthiness

- Chairmanship of CEN-CENELEC JTC21
- Steering group for German AI Standardization Roadmap
- Convenorship of IEC SEG 10 AI Ethics
  AI Trust Standard & Label (April 2022)
- Observer in Council of Europe Committee on AI
- Lead & Conception "AI Ethics: From Principles to Practice" (April 2020)
- OECD ONE.AI Co-Chair Classification and risk assessment
- German Enquete on AI: Advice to MPs; Federal Ministry of Labour and Social affairs: Framework for AI Training Data Quality
- Advice to EU-Commission DG GROW / DG JUST / DG CNECT

Broader perspective: from AI Ethics to AI quality

AI Quality Summit, 2 November 2022 near Frankfurt Airport

Registration: www.ai-q.de

The big challenge

Operationalise AI Ethics with an approach …

- … that is viable for industry, regulators and consumers / citizens
- … and that makes ethics measurable and enforceable

Why standardisation is the right approach

Standardisation =

1. Building consensus among all relevant stakeholders
2. Formulating this consensus in a concrete, specific, practically useful way

How to handle AI Ethics through standardization

Explicit ethical rules (e.g., "Child more important than old person", "100 severely injured better than 1 dead"): consensus unlikely

Standardised description of ethical aspects of systems (e.g., Privacy A, Transparency D, Fairness B): viable, flexible and strong

Only processes and structures for decisions about ethics (e.g. ethics boards in companies): viable but limited on its own

Approach: A standardised "label" / "short datasheet" that can be attached to AI products

Classification of aspects such as transparency, accountability, privacy, fairness, reliability

- provides **positive differentiation** in the marketplace
- ensures **fair competition**
- promotes consistency with **organisational and societal values**
- facilitates **compliance** with regulation
- supports policymakers in **minimising red tape**

European and international standardization

- CEN-CENELEC Focus Group for Artificial Intelligence
- CENELEC Roadmap report October 2020
- IEC SEG 10 Ethics in autonomous and artificial intelligence applications
  Final report July 2021

Ai Ethics Impact Group www.ai-ethics-impact.org

- Bertelsmann Stiftung
- Tübingen University
- International Center for Ethics in the Sciences and Humanities (IZEW)
- ITAS Institut für Technikfolgenabschätzung und Systemanalyse
- HLRS High-Performance Computing Center Stuttgart
- Technical University Kaiserslautern
- iRights.Lab Think tank for the digital world

Comprehensive consortial standard 2021/22

Version 1 published in April 2022

Describes the characteristics of an AI product with regards to transparency, accountability, privacy, fairness, reliability

Question: Which categories do we include?

- transparency
- justice
- accountability
- privacy
- reliability/safety
- environmental sustainability

Question: How can we measure transparency, accountability, etc.?

Each indicator can correspond to one of three categories, which are hierarchically subordinate:

Negative anchor indicator:

*"Necessary condition"; p*rerequisite for T1.2 and T1.3., minimum requirement (e.g. E-G)

Positive anchor indicator:

"Sufficient condition"; The fulfilment of one indicator can substitute the fulfilment of one or more other indicators.

Score indicators:

Build on anchor indicators.

Scoring of the score indicators are added and averaged to determine the level of the label

Question: What levels are acceptable in a given application?

AI Trust Standard & Label from a company perspective

Ai Trust Alliance

Company

1. Information, Consultation
2. Membership and Licence: By becoming a member of the "AI Trust Alliance", the company agrees to a licence agreement. Members are authorised to use the label for qualified products.
3. Product registration: The product to be tested is registered.
    a. External audit
       Company enters a contract with an approved certification institution. The certification institution tests/audits according to the criteria of the VDE SPEC and provides a certificate.
    b. Self- declaration
       Self-declaration of compliance with VDE SPEC based on in-house evaluation
4. Usage/Publication
   After passing the product qualification, the AI Trust Label can be used. The product is listed in the AI Trust product database. For many products concurrent conformity according to Art. 40/41 AI Act.

Towards a European approach

Combining complementary work (Confiance AI – EU – CEN/CENELEC)

Metrics, tools, governance

Use results for standardization and regulation (CEN, CENELEC, JTC21, EU)