KI-FACHKONFERENZ

AI Audit

Martin
Saerbeck

22.11.2021

AI Qualitätsmanagement

- Was wird erwartet? Welche Unterstützung gibt es?
- Welche industrieübergreifenden Anforderungen benötigen wir?

KI Herausforderungen

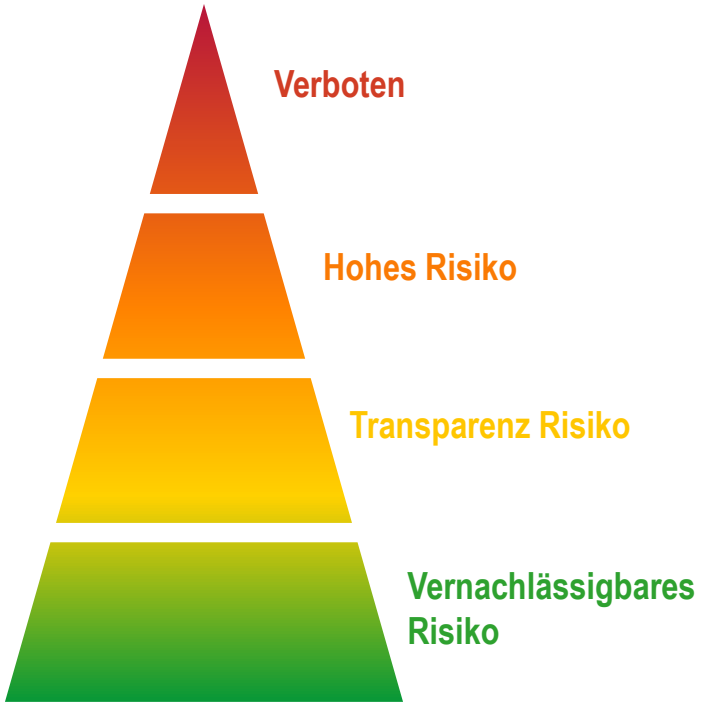
- KI implementiert eine Näherung zu einer unbekannten Zielfunktion
- Daten übernehmen die Roller der Spezifikation (zusammen mit dem Optimierungsziel)
- Prozess beinhaltet statistische Elemente (z.B. Reihenfolge der Beispiele, Initialisierung, statistisches Rauschen)
- Identifikation von lokalen und globalen Minima
- Hoch-dimensionale Daten- und Parameterräume
- Balance zwischen sich ausschließenden Qualitätsmerkmalen (z.B. Genauigkeit vs. Bias)
- Nicht-linearität zwischen Eingabe und Ausgabe (kleine Änderungen in der Eingabe können große Auswirkungen haben)
- Entwicklungsprozess ist derzeit nicht hinreichend formalisiert (z.B. Dokumentation, wann ist das Lernen abgeschlossen?)
- Feld aktiver Forschung (neue Erkenntnisse definieren bewährte Vorgehensweise ständig neu
- Fehlende Ausbildung, Erfahrung, Hilfsmittel, Zeit der Experten
- Wettbewerb

AI Act

- Anhang 3 listet Anwendungsbereiche mit hohem Risiko

- Anforderungen an:
  - Datenqualität
  - Transparenz
  - Menschliche Aufsicht
  - …

- Qualitätsmanagement Anforderungen
- Post-Market Monitoring



Standards (nicht-KI)

Text	Text	Text
Life Cycle	ISO/IEC/IEEE 12207:2017	Systems and software engineering — Software life cycle processes
	ISO/IEC/IEEE 15288:2015	Systems and software engineering — System life cycle processes
Quality	ISO 25000 Family	Systems and software engineering — Systems and software quality requirements and evaluation
	ISO 9000 Family	Quality management systems
Testing	ISO/IEC/IEEE 29119	Software and systems engineering – Software testing
	ISO/IEC 33002	Information technology — Process assessment — Requirements for performing process assessment
Safety	IEC 61508	Functional safety of electrical/electronic/programmable electronic safety-related systems
	ISO 26262:2018	Road vehicles — Functional safety
Security	ISO/IEC 27000 Family	Information Security
	IEC 62443	Industrial communication networks – Network and system security
Risk	ISO 12100:2010	Safety of machinery — General principles for design — Risk assessment and risk reduction
	ISO 31000:2018	Risk Management
Assessment	ISO/IEC 33002:2015	Information technology — Process assessment — Requirements for performing process assessment

Technische Standards – ISO SC42

WG 1: Foundational Standards

- **ISO/IEC 22989** Artificial Intelligence –Concepts and Terminology
- **ISO/IEC 23053** Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
- **ISO/IEC 42001** Artificial Intelligence –Management System

WG 5: Computational Approaches and Computational Characteristics of AI Systems

- **ISO/IEC 24372** Information technology — Artificial intelligence (AI) — Overview of computational approaches for AI systems
- **ISO/IEC TS 4213** Information technology — Artificial intelligence — Assessment of Machine Learning Classification Performance
- **ISO/IEC WD 5392** Information technology — Artificial intelligence — Reference Architecture of Knowledge Engineering

JWG 1: IT - Governance of AI

- **ISO/IEC 38507** Information Technology —Governance of IT — Governance implications of the use of artificial intelligence by organizations

WG 2: Data

- **ISO/IEC 24668 Process management framework for big data analytics**
- **ISO/IEC 5259-x Data quality for analytics and ML**
  - **5259-1** – Overview, terminology, and examples
  - **5259-2** – Data quality measures
  - **5259-3** – Data quality management requirements and guidelines
  - **5259-4** – Data quality process framework
  - Part 5: Data quality assurance
  - Part 6: Data quality governance
  - Part 7: Data quality visualization

- **ISO/IEC 20546:2019** Information technology — Big data — Overview and vocabulary
- **ISO/IEC TR 20547-1:2020** Information technology — Big data reference architecture — Part 1: Framework and application process
- **ISO/IEC TR 20547-2:2018** Information technology — Big data reference architecture — Part 2: Use cases and derived requirements
- **ISO/IEC 20547-3:2020** Information technology — Big data reference architecture — Part 3: Reference architecture
- **ISO/IEC TR 20547-5:2018** Information technology — Big data reference architecture — Part 5: Standards roadmap

WG 3: Trustworthiness

- **ISO/IEC 24027** Bias in AI systems and AI aided decision making
- **ISO/IEC 24029-2** Assessment of the robustness of neural networks - Formal methods methodology
- **ISO/IEC 23894** Risk Management
- **ISO/IEC 24368** Overview of ethical and societal concerns
- **ISO/IEC 5469** Functional safety and AI systems
- **ISO/IEC 25059** (SQuaRE) – Quality model for AI systems
- **ISO/IEC 6254** Explainable AI
- [NPTS] 5471 Quality evaluation guidelines for AI systems

- **24028:2020** Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence
- **ISO/IEC 24029-1:2021** Assessment of the robustness of neural networks – Overview

WG 4: Use cases and applications

- **TR 24030** Artificial Intelligence: Use cases
- **5338** Artificial intelligence - AI system life cycle processes
- **5339** Artificial Intelligence - Guidelines for AI Applications

Datenqualitätsstandards

- ISO 8000 Data quality
- ISO/IEC 25012 SQuaRE Data quality model; ISO/IEC 25024 SQuaRE Measurement of data quality
- ISO/IEC 5259 Data quality for analytics and machine learning
- ISO/IEC 24668 Process management framework for big data analytics

Data Quality

- Accuracy
- Precision
- Completeness
- Consistency
- Relevance
- Scalability
- Portability
- Timeliness
- Identifiability
- Auditability
- Credibility
- Accessibility
- Confidentiality
- Efficiency
- Ethical alignment
- Understandability
- Recoverability
- Portability

Vielen Dank für Ihre Aufmerksamkeit!

Ihr Ansprechpartner:

Martin Saerbeck
TÜV SÜD - CTO Digital Service

Tel.: +65 96777026
E-Mail: martin.saerbeck@tuvsud.com