Internationale AI Standardisierung – Rahmen und Trustworthiness im Fokus

@Plattform Industrie 4.0
22.10.2021

Martina Paul, MBA & Rania Wazir, PhD

Vorsitz und Stv Vorsitz ASI AG 001.42

Artificial Intelligence

Freier Welthandel - Rahmen

WORLD TRADE ORGANIZATION

TBT Agreement

(Agreement on Technical Barriers to Trade)

GPA Agreement

(Agreement on Government Procurement)*

*an appendix of the TBT Agreement, requires countries party to the agreement to define a technical specification based on the applicable international standard (if one exists) when they carry out government procurement that exceeds a certain size.

WTO TBT Agreement

Voluntary	Mandatory
Standards	Technical Regulations
Conformity Assessment	

Use of relevant international standards

International Standards - Prinzipien

- ✓ Transparency
- ✓ Openness
- ✓ Impartiality & Consensus
- ✓ Effectiveness & Relevance
- ✓ Coherence
- ✓ Development Dimension

Die Garanten des freien Handels

WORLD TRADE ORGANIZATION

World Standards Cooperation

Non-Governmental Organization

Products, Manufacturing

Stakeholders: mainly industry

Non-Governmental Organization

Processes, Architectures, Society

Stakeholders: all those who have a legitimate interest

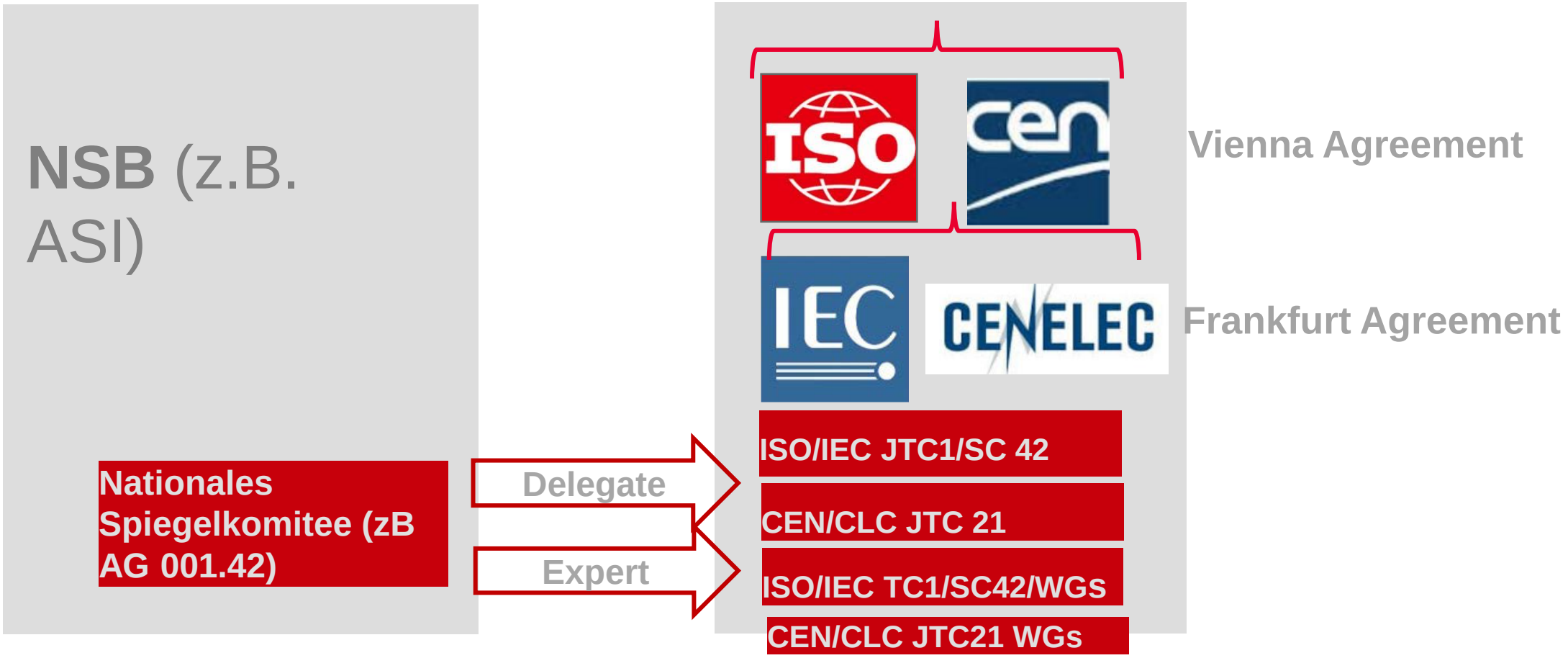
Intergovernmental Organization

(Specialised Agency of the UN)

Telecommunication

Stakeholders: Regulators, Industry

Die Bedeutung der nationalen Standardisierung



Trustworthy AI & Standards

Verantwortung & Vertrauen in einer algorithmusgetriebenen Welt

Künstliche Intelligenz ist ...

... wenn ein Algorithmus Aufgaben erledigt, die wir normalerweise einer menschlichen Intelligenz zuordnen (Text verstehen, Bilder erkennen, Entscheidungen treffen)

Wenn der Algorithmus...

1. ... aus Schritt für Schritt Anweisungen besteht
   nennt man das symbolic AI

2. ... anhand von vielen Beispielen lernt
   nennt man das Maschinelles Lernen
   Deep Learning ist nur eine von vielen möglichen Methoden

Künstliche Intelligenz ist ...

... weder objektiv, noch neutral

Welche Anwendungen werden entwickelt:

Vincent,J., ***Deepfake bots on Telegram make the work of creating fake nudes dangerously easy***, the Verge, Oktober 2020, https://www.theverge.com/2020/10/20/21519322/deepfake-fake-nudes-telegram-bot-deepnude-sensity-report

Welche Funktionalitäten sollen sie haben:

Duhaime-Ross, A., ***Apple promised an expansive health app, so why can't I track menstruation?***, the Verge, September 2014, https://www.theverge.com/2014/9/25/6844021/apple-promised-an-expansive-health-app-so-why-cant-i-track

Wie und was wird klassifiziert:

Leufer, D., ***Computers are binary, people are not: how AI systems undermine LGBTQ identity*****,** Access Now, April 2021, https://www.accessnow.org/how-ai-systems-undermine-lgbtq-identity/

Trustworthy?

Transparency
Explainability
Accountability
Fairness & Non-discrimination
Privacy
Robustness
Safety

Trustworthy? Wer definiert?

Transparency
Explainability
Accountability
Fairness & Non-discrimination
Privacy
Robustness
Safety

STANDARDS

RAHMEN Artificial Intelligence Standardisierung

- ✓EC Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL (N0186)
- ✓EU Whitepaper Artificial Intelligence (N0094)
- ✓EU Single Digital Gateway Regulation (N100) und Architecture Vision
- ✓EU Data Strategy (N0093)
- ✓AIM 2030 – Österreichische KI Strategie (N0069)
- ✓EU Ethic Guidelines for Trustworthy AI (N0041)

Facts & Figures

- Spiegelkomitee zu drei JTC1 Subcommittees und einem CEN/CLC JTC
- 51 teilnehmende Personen
- Aktive Kommentierung und Beiträge zu sechs Entwürfen bzw Ballots
- 17 ExpertInnen in WGs (SC 42 & SC 22, CEN/CLC JTC21 WG1, AHGs)
- Regelmäßige Vorträge zu Standardisierungsthemen (GAIA X, ÖCloud, AI Standardisierungsroadmap Dtl).

ECOSYSTEM

OVERALL MORE THAN 60 LIAISONS

50 Members 33 P-Members/17 O-Members

STRUKTUR

ISO/IEC JTC 1/SC 42/AG 2	**AI Systems Engineering**
ISO/IEC JTC 1/SC 42/AHG 1	**Dissemination and outreach**
ISO/IEC JTC 1/SC 42/AHG 2	**Liaison with SC 38**
ISO/IEC JTC 1/SC 42/AHG 4	**Liaison with SC 27**
ISO/IEC JTC 1/SC 42/AHG 5	**AI standardization landscape and roadmap**
ISO/IEC JTC 1/SC 42/JWG 1	**Joint Working Group ISO/IEC JTC1/SC 42 - ISO/IEC JTC1/SC 40: Governance implications of AI**
ISO/IEC JTC 1/SC 42/WG 1	**Foundational standards**
ISO/IEC JTC 1/SC 42/WG 2	**Data**(+ AHG AI Data und AHG Data Quality)
ISO/IEC JTC 1/SC 42/WG 3	**Trustworthiness**
ISO/IEC JTC 1/SC 42/WG 4	**Use cases and applications**
ISO/IEC JTC 1/SC 42/WG 5	**Computational approaches and computational characteristics of AI systems**

CEN/CLC JTC21 – WGs

- ✓ WG1: Strategic Advisory Group
- ✓ AHG1: Augmented Goal Specification
- ✓ AHG2: AI Conformity Assessment
- ✓ AHG3: Green
- ✓ AHG4: Speech Recognition

PUBLISHED STANDARDS

ISO/IEC TR 24028:2020
Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence

ISO/IEC TR 20547-5:2018
Information technology — Big data reference architecture — Part 5: Standards roadmap

ISO/IEC 20547-3:2020
Information technology — Big data reference architecture — Part 3: Reference architecture

ISO/IEC TR 20547-2:2018
Information technology — Big data reference architecture — Part 2: Use cases and derived requirements

ISO/IEC TR 20547-1:2020
Information technology — Big data reference architecture — Part 1: Framework and application process

ISO/IEC 20546:2019
Information technology — Big data — Overview and vocabulary

ISO/IEC TR 24029-1:2021
Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview

ISO/IEC TR 24030:2021
Information technology — Artificial intelligence (AI) — Use cases

Diskriminierung, Vorurteile, Bias ...

- Simonite, T., ***When It Comes to Gorillas, Google Photos Remains Blind,*** Wired, November, 2018 https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/

- Ziad Obermeyer et al. ***Dissecting racial bias in an algorithm used to manage the health of populations***. In: Science 366.6464 (2019), pp. 447–453. https://science.sciencemag.org/content/366/6464/447

- Geiger, G., ***Court Rules Deliveroo Used 'Discriminatory' Algorithm*****,** Motherboard, January 5 (2021), *https://www.vice.com/en/article/7k9e4e/court-rules-deliveroo-used-discriminatory-algorithm*

- The Guardian, ***Amazon ditched AI recruiting tool that favored men for technical jobs,*** Reuters, Oktober, 2018 https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

KI Ethik Richtlinien

- **OECD** https://www.oecd.ai/ai-principles
- **UNESCO:** https://en.unesco.org/news/unesco-launches-worldwide-online-public-consultation-ethics-artificial-intelligence
- **UNICEF:** https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children
- **EU:** https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en
- **EU:** https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment
- **Microsoft AETHER:** https://www.microsoft.com/en-us/ai/responsible-ai
- **Google**'s (former) Ethics Committee: https://www.reuters.com/article/us-alphabet-google-ai-idUSKCN1RH00S
- **Partnership on AI:** *https://www.partnershiponai.org/*

Mit Stand 2019*, gibt es über 84 KI Ethik Richtlinien.

Seien Sie Bias Detektiv!

Dieses Experiment dürfen Sie gerne zu Hause ausführen!*

Geben Sie folgende Texte in Google Translate rein, und übersetzen Sie aus dem Englischen, ins Deutsche:
Englisch: My teacher is clever. He immediately found the solution
Google Deutsch:

Englisch: My teacher is clever. She immediately found the solution
Google Deutsch:

Englisch: My doctor is clever. She immediately found the solution
Google Deutsch:

***Hat tip Liad Magen für die Idee zum Experiment.**

FRAGEN?

DANKE FÜR IHRE AUFMERKSAMKEIT!

Martina Paul, MBA
www.linkedin.com/in/martinapaul1

Rania Wazir, Ph.D.
www.linkedin.com/in/raniawazir