KI-FACHKONFERENZ

Workshop: Transparenz & Erklärbarkeit

Tanja Hagemann, Deutsche Telekom AG
Benjamin Ledwon, Deutsche Telekom AG
Taras Holoyad, Bundesnetzagentur

Berlin, 22.11.2021

Agenda

- Intro: EU Artifical Intelligence Act
- Transparenz und Erklärbarkeit von KI
- KI Standardisierung
- Diskussion

  → **Voting**

Intro
EU Artifical Intelligence Act

Quelle: ...

Intro: EU Artifical Intelligence Act

KI Definition*

“

Software that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with, and that is developed with one or more of the following techniques and approaches:

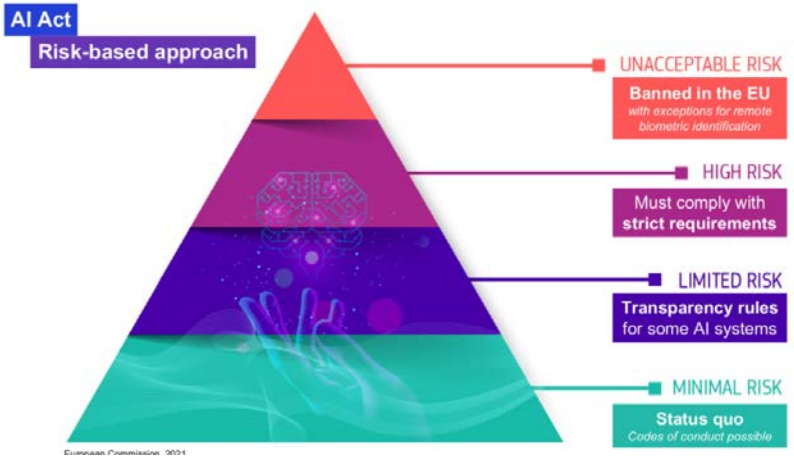
- **Machine learning** approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
- **Logic- and knowledge-based** approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
- **Statistical approaches**, **Bayesian** estimation, **search** and **optimization** methods.

”

* From the EU Artificial Intelligence Act.
This definition has received a lot of comments that state that it is too broad and would include various traditional software solutions like PLC programs which are already sufficiently cover by existing regulation.

Intro: EU Artifical Intelligence Act

- **Ziel des Gesetzesvorschlags**: Marktzugang von KI-Systemen in den europäischen Binnenmarkt auf Grundlage klarer technischer Anforderungen.
- **Struktur des Vorschlags**: Definition von KI, Risikoklassifizierung (niedrig, hoch, Komplettverbot), technische Anforderungen, Konformitätsprüfung & Marktbeobachtung, Governance.
- **Timeline**: Verhandlungen im Rat am weitesten fortgeschritten, Europäisches Parlament steht noch am Anfang. Verabschiedung 2023 möglich.
- **Anwendungsbeispiele** & **Relevanz** am Beispiel der Deutschen Telekom

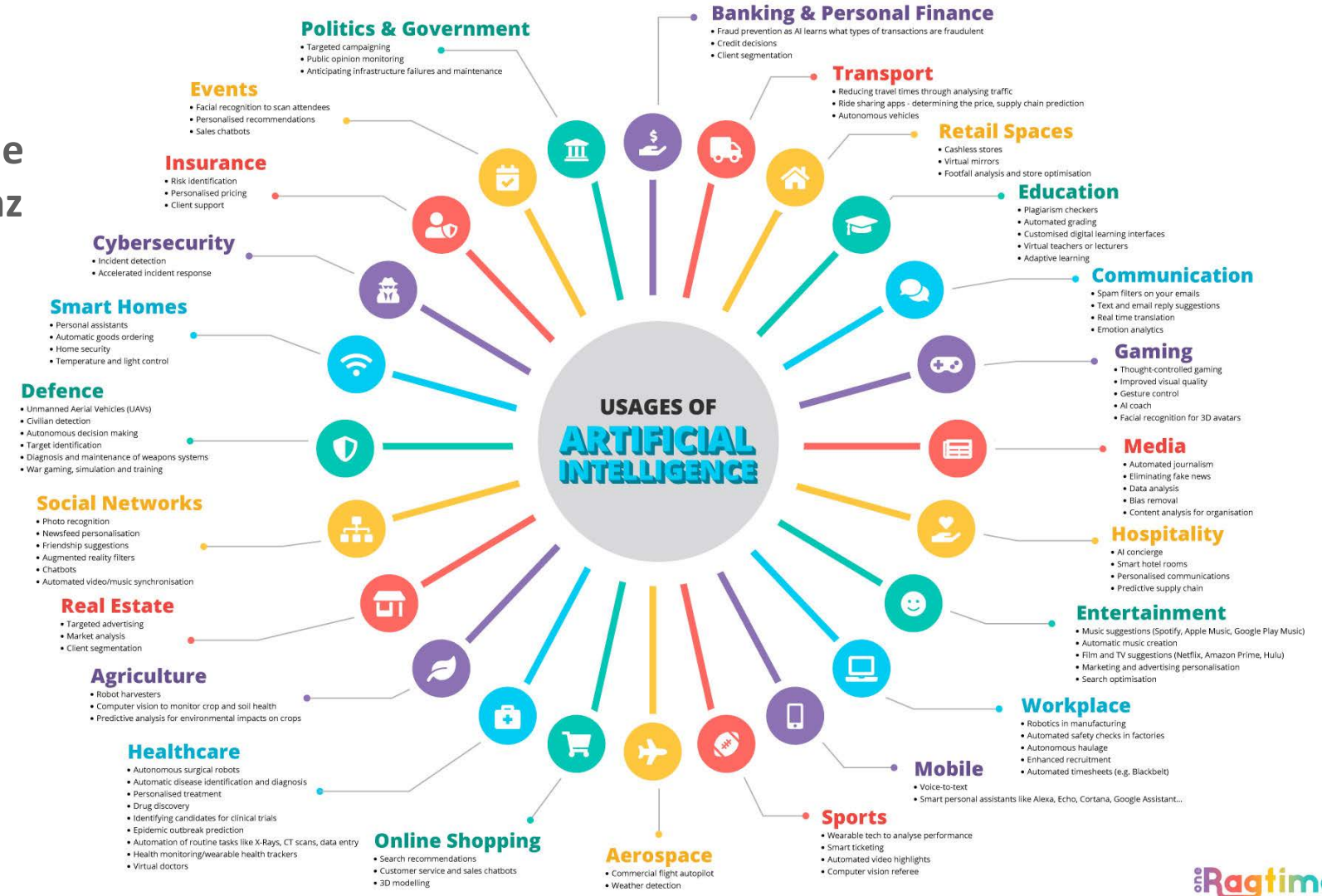


Welche industrieübergreifenden Anforderungen an **Transparenz & Erklärbarkeit für KI** benötigen wir?

Transparenz & Erklärbarkeit von KI

Quelle: ...

Künstliche Intelligenz in der Industrie



Transparenz vs. Erklärbarkeit

Definition

“

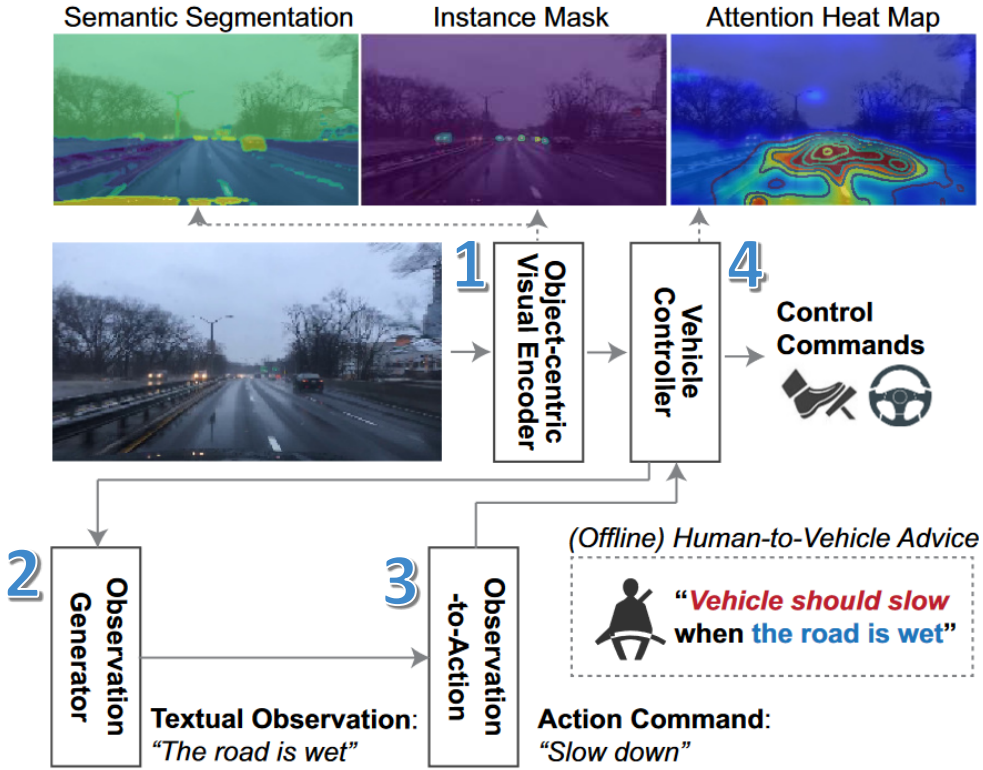
Liegt **Transparenz** eines KI-Modells vor, so kann es unter der Voraussetzung nachvollziehbarer Eingangsdaten auch als „**White-Box**“-Modell bezeichnet werden. Bei der **Erklärbarkeit** geht es hingegen darum, einer Zielperson eine verständliche **Begründung** aktiv bereitzustellen, die es ihr ermöglicht, das Ergebnis eines KI-Modells nachzuvollziehen.

”

Quelle: ERKLÄRBARE KI - Anforderungen, Anwendungsfälle und Lösungen im Auftrag des Bundesministeriums für Wirtschaft und Energie

Erklärbare KI
Beispiel: Autonomes Fahren

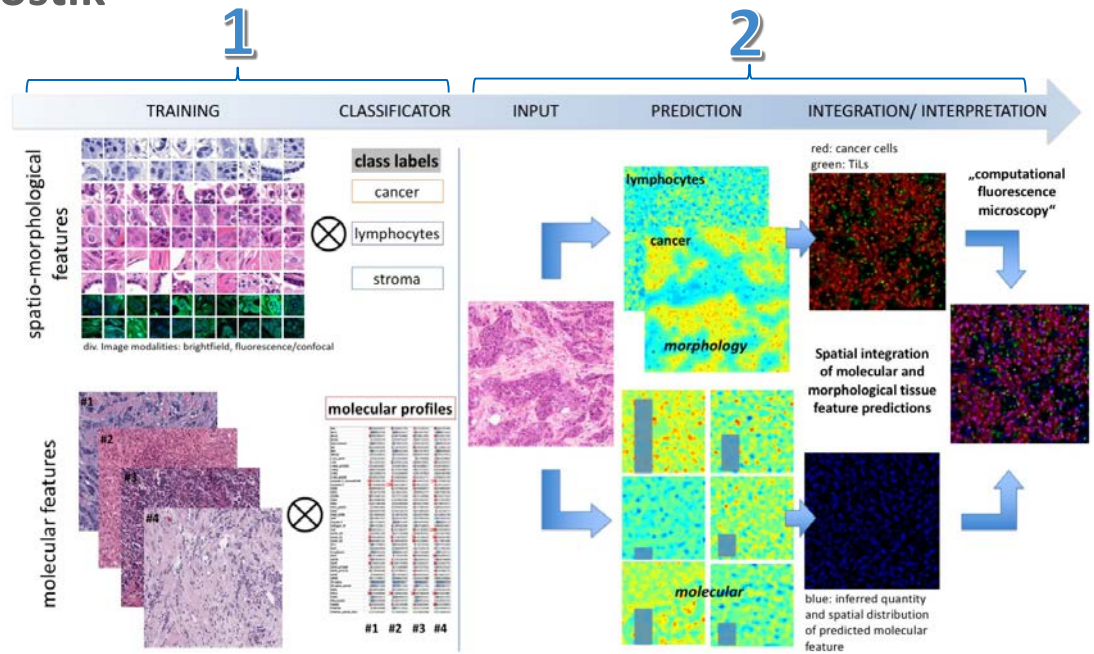
1. Objektzentrierter visueller Encoder, der auf einem semantischen Segmentierungsmodell aufbaut
2. Beobachtungsgenerator, der textuelle Beobachtungen über die Szenen erzeugt ("Die Straße ist nass")
3. Modul zur Umwandlung von Beobachtungen in Aktionen, das eine visuelle Szenenbeschreibung auf einen (high-level) Aktionsbefehl abbildet ("Langsamer fahren")
4. Fahrzeugsteuerung auf der Grundlage des generierten Handlungsbefehls



Kim, Jinkyu, et al. "Advisable learning for self-driving vehicles by internalizing observation-to-action rules." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

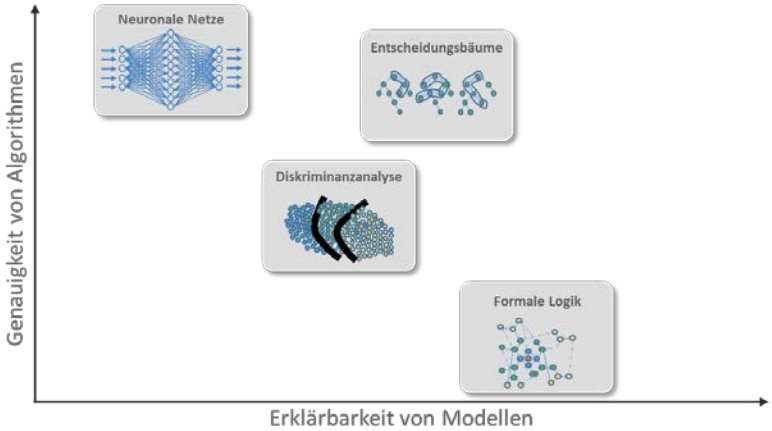
Erklärbare KI
Beispiel: Medizinische Diagnostik

1. Morphologische, molekulare und histologische Daten werden in eine einzige Analyse integriert
2. Das System liefert eine Verdeutlichung des KI-Entscheidungsprozesses in Form von Heatmaps. Diese Heatmaps zeigen Pixel für Pixel, welche visuellen Informationen den KI-Entscheidungsprozess in welchem Ausmaß beeinflusst haben, so dass Ärzte die Ergebnisse der KI-Analyse verstehen und deren Plausibilität beurteilen können.



Binder, Alexander, et al. "Morphological and molecular breast cancer profiling through explainable machine learning." *Nature Machine Intelligence* 3.4 (2021): 355-366.

Transparenz & Erklärbarkeit: KI-Methoden



A. Barredo Arrieta et al.:
Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI

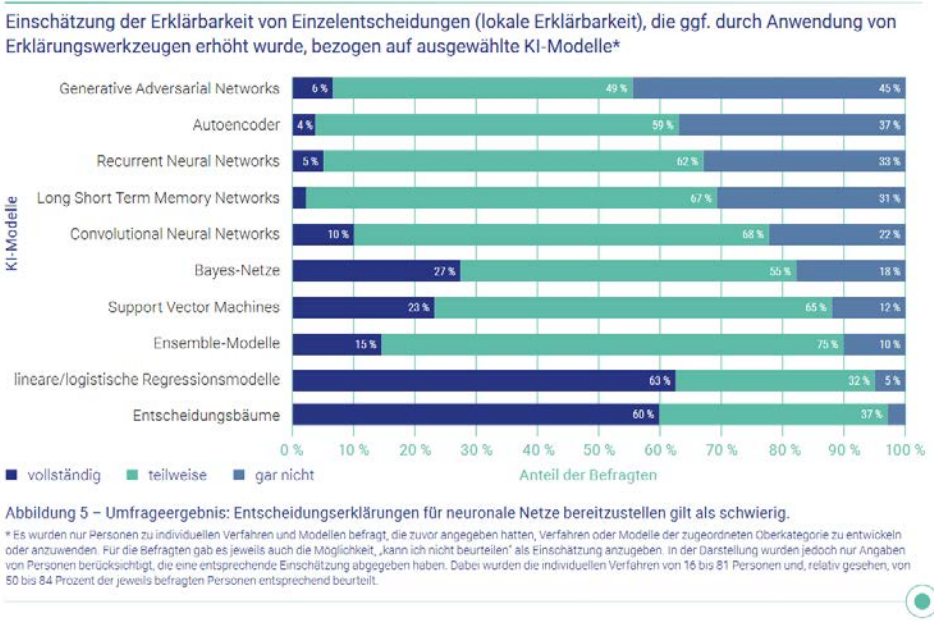


Abbildung 5 – Umfrageergebnis: Entscheidungserklärungen für neuronale Netze bereitzustellen gilt als schwierig.

* Es wurden nur Personen zu individuellen Verfahren und Modellen befragt, die zuvor angegeben hatten, Verfahren oder Modelle der zugeordneten Oberkategorie zu entwickeln oder anzuwenden. Für die Befragten gab es jeweils auch die Möglichkeit, „kann ich nicht beurteilen" als Einschätzung anzugeben. In der Darstellung wurden jedoch nur Angaben von Personen berücksichtigt, die eine entsprechende Einschätzung abgegeben haben. Dabei wurden die individuellen Verfahren von 16 bis 81 Personen und, relativ gesehen, von 50 bis 84 Prozent der jeweils befragten Personen entsprechend beurteilt.

Quelle: ERKLÄRBARE KI - Anforderungen, Anwendungsfälle und Lösungen im Auftrag des Bundesministeriums für Wirtschaft und Energie

Transparenz & Erklärbarkeit: Domänen und Nutzergruppen

Bedeutung Erklärbarkeit von Einzelentscheidungen (lokale Erklärbarkeit) nach Anwendungsbranchen*

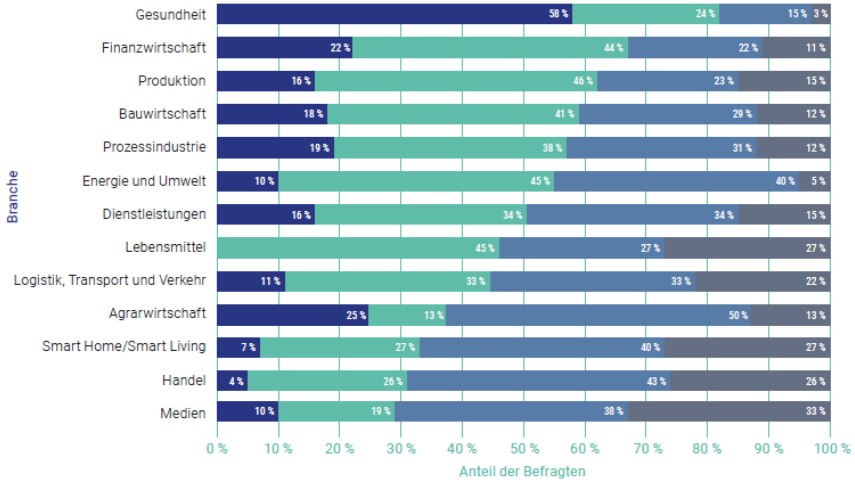


Abbildung 6 – Umfrageergebnis: lokale Erklärbarkeit in Branchen Gesundheit, Finanzwirtschaft, Produktion am stärksten gefordert.

* Es wurden nur Personen zu einzelnen Anwendungsbranchen befragt, die zuvor angegeben hatten, in der jeweiligen bzw. für die jeweilige Anwendungsbranche KI-Systeme anzuwenden bzw. zu entwickeln. Die Branchenreihenfolge in der Abbildung wurde gemäß der empfundenen Bedeutung von Entscheidungserklärungen bzw. lokaler Erklärbarkeit sortiert; hier die Kategorien „zwingend erforderlich" oder „sehr wünschenswert". Bei sonstigen Anwendungsbranchen, die von mehreren Personen angegeben wurden, wurde vor allem für das vergleichsweise unspezifische Anwendungsgebiet „IT / Software" die lokale Erklärbarkeit mehrfach als sehr wünschenswert oder zwingend erforderlich eingestuft. Für weitere einzelne Ergänzungen wie „Legal Tech", Personalwesen und Öffentliche Sicherheit, die auch unter „Sonstiges" eingeordnet wurden, galt die lokale Erklärbarkeit für die betreffenden Personen als zwingend erforderlich.

Einschätzung der Bedeutung von Entscheidungserklärungen für Zielgruppen*

Zielgruppe	Erklärbarkeit von Einzelentscheidungen (lokale Erklärbarkeit)			
	heute	Zukunft (5–10 Jahre)	Trend	
KI-Entwickler:innen	76 %	56 %	▼	-20 %
Domänenexpert:innen	59 %	59 %	▶	-1 %
Die Management-Ebene	38 %	57 %	▲	19 %
Endkund:innen, Endnutzer:innen	35 %	65 %	▲	31 %
Interne Prüfer:innen	41 %	57 %	▲	16 %
Externe Prüfer:innen	35 %	63 %	▲	28 %

Abbildung 7 – Umfrageergebnis: Erklärbarkeit heute besonders für KI- und Domänenexpert:innen wichtig, künftig wird eine vergleichbare Bedeutung für fast alle Zielgruppen prognostiziert.

* Aus der Grafik ist ablesbar, wieviel Prozent der Umfrageteilnehmer:innen die jeweilige Gruppe als bedeutende Zielgruppe für lokale Erklärbarkeit einschätzen.

Quelle: ERKLÄRBARE KI - Anforderungen, Anwendungsfälle und Lösungen im Auftrag des Bundesministeriums für Wirtschaft und Energie

KI Standardisierung

Quelle: ...

Weltweite KI - Standardisierung (1/2)

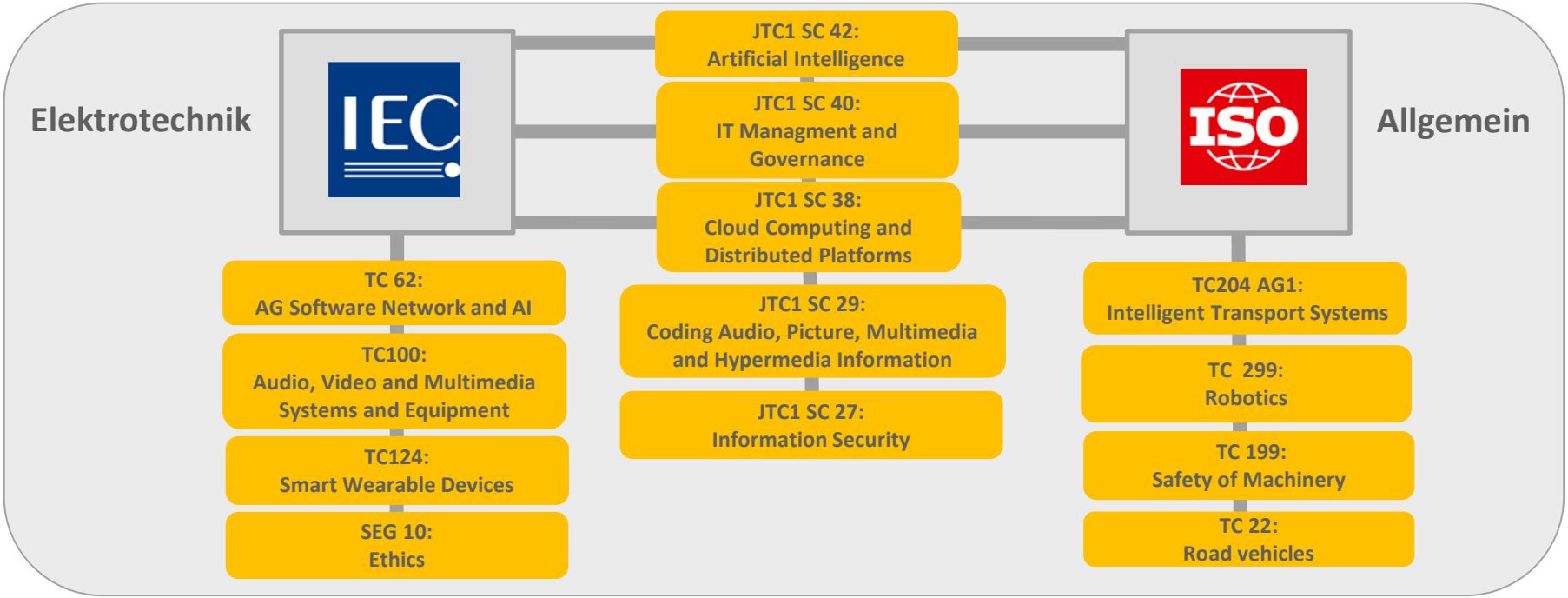
ITU: International Telecommunication Union (Internationale Fernmeldeunion)

Focus Groups:

- Environmental Efficiency for AI and other Emerging Technologies
- AI for autonomous and assisted driving
- Autonomous Networks
- Artificial Intelligence (AI) and Internet of Things (IoT) for Digital Agriculture
- AI for Natural Disaster Management

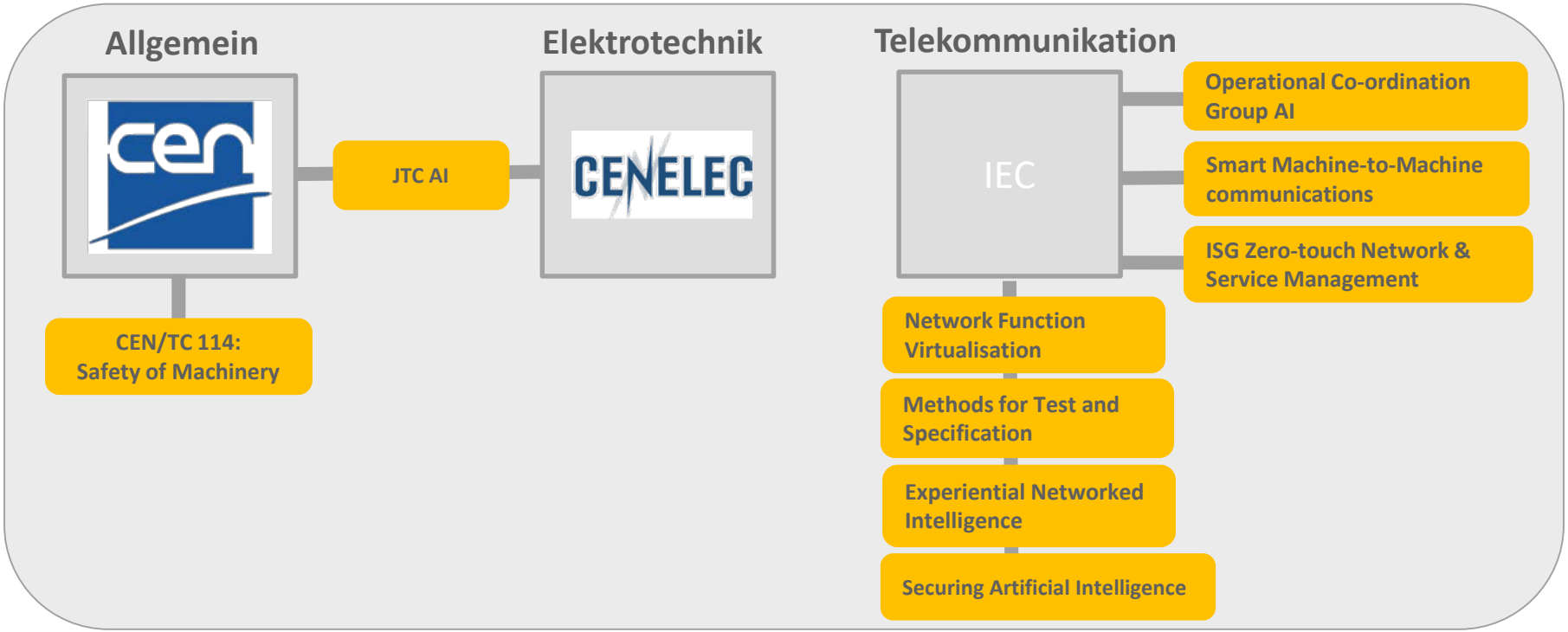
ITU-T Focus Groups: https://www.itu.int/en/ITU-T/focusgroups/Pages/default.aspx

Weltweite KI - Standardisierung (2/2)



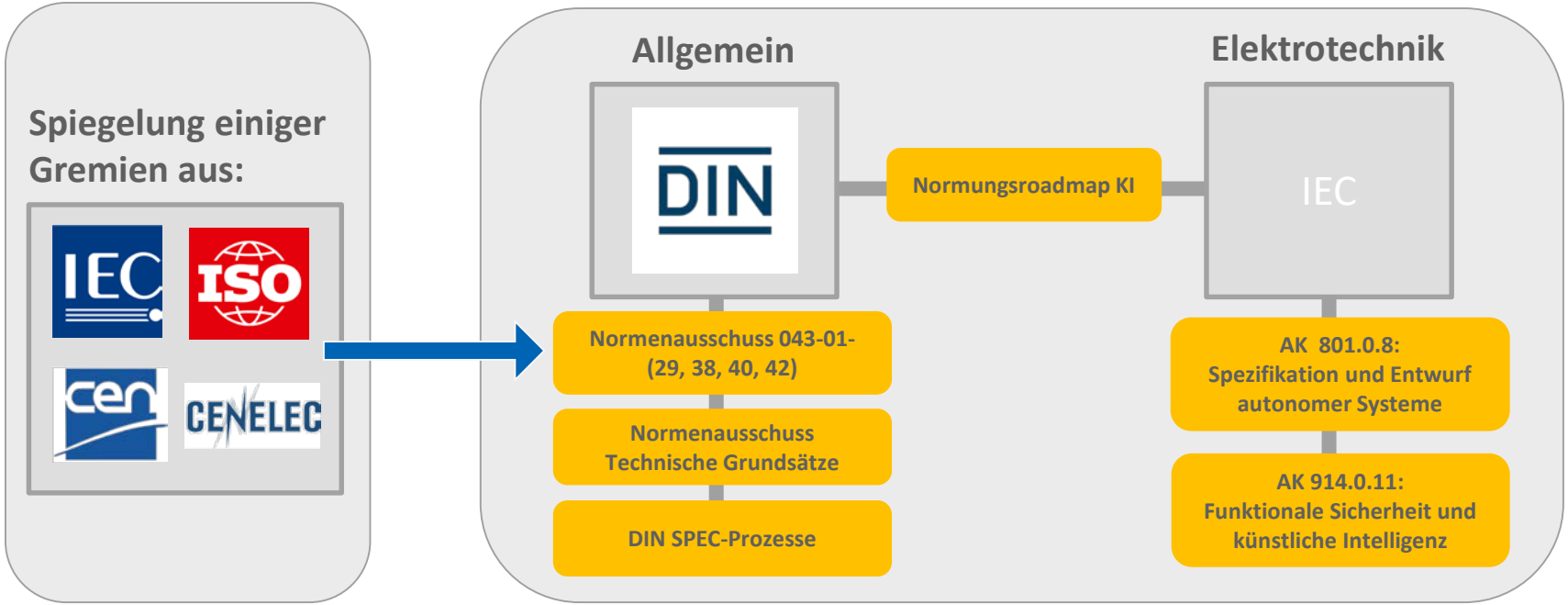
ISO/IEC JTC 1 Information technology: https://www.iso.org/committee/45020.html
Harmonisierung durch Normen und Standards: https://www.sci40.com/sci-4-0/go-global/

KI - Standardisierung in Europa



Europäische Standardisierung: https://www.cencenelec.eu/european-standardization/

KI - Standardisierung in Deutschland



Harmonisierung durch Normen und Standards: https://www.sci40.com/sci-4-0/go-global/

AI-Normen in ISO/IEC JTC1 SC42 mit Bezug auf die EU AI-Verordnung

ISO/IEC hat einige genehmigte Dokumente und viele laufende Arbeiten, die in irgendeiner Weise mit der EU AI-Verordnung zu tun haben. Im Folgenden sind nur einige davon aufgeführt. Es ist eine detaillierte Analyse erforderlich, inwieweit sie den Anforderungen der Verordnung entsprechen.

General:

- ISO/IEC 22989 Artificial Intelligence Concepts and Terminology

Risk management:

- ISO/IEC 23894 AI Risk management

Data Quality:

- ISO/IEC IS 5259 series Data quality for analytics and machine learning (ML)

Transparency:

- ISO/IEC TS 6254 Objectives and approaches for explainability of ML models and AI systems
- New work item proposed Transparency taxonomy of AI systems

Human oversight:

- ISO/IEC TS 8200 Controllability of automated artificial intelligence systems

Robustness, accuracy and cybersecurity:

- ISO/IEC 24029 series Assessment of the Robustness of Neural Networks
- ISO/IEC TS 4213 Assessment of machine learning classification performance

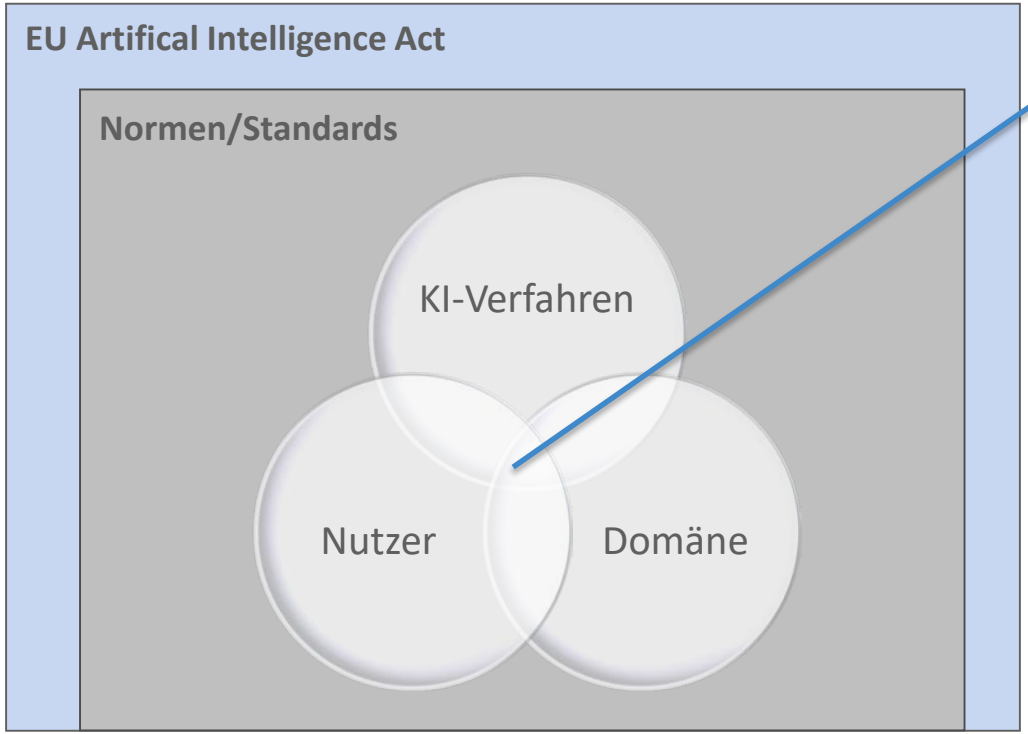
Quality management:

- ISO/IEC 42001 AI management system

Functional safety:

- ISO/IEC TR 5469 Functional Safety and AI systems

Transparenz & Erklärbarkeit für KI



Was denken Sie?

Bei der Verwendung welcher KI-Verfahren besteht in Ihrer Domäne (zukünftig) ein Bedarf an Normen/Standards für Transparenz/Erklärbarkeit? Für welche Nutzergruppe/n?

Link zum Voting:
https://vote.telekom.net/464088

Passwort:
DINKI2021

Voting

Link zum Voting: https://vote.telekom.net/464088
Passwort: DINKI2021

Frage 1	In welcher Industrie sind Sie tätig?
Frage 2	In welche Risiko-Bereiche fallen Ihrer Meinung nach die KI-Anwendungen in Ihrer Industrie?
Frage 3	Besteht Ihrer Meinung nach in Ihrer Domäne Bedarf an Transparenz/Erklärbarkeit für KI? Wenn ja, für welche KI-Verfahren und welche Nutzergruppe/n?
Frage 4	Im Hinblick auf den EU Artificial Intelligence Act: Kennen Sie eine/n bereits bestehende/n KI-Norm/Standard, die/den Sie uns mitteilen wollen?
Frage 5	Im Hinblick auf den EU Artificial Intelligence Act: Kennen Sie eine/n sich in der Entwicklung befindliche/n KI-Norm/Standard, die/den Sie uns mitteilen wollen?
Frage 6	Im Hinblick auf den EU Artificial Intelligence Act: Welche speziellen Normen für KI-Transparenz/Erklärbarkeit braucht es aus Ihrer Sicht zukünftig noch?
Frage 7	Welches Feedback möchten Sie uns noch mit auf den Weg geben?

Diskussion

Quelle: ...

Vielen Dank für Ihre Aufmerksamkeit!

Ihr Ansprechpartner:

Tanja Hagemann
Telekom Innovation Laboratories

linkedin.com/in/tanjahagemann
E-Mail:
Tanja.Hagemann@telekom.de

Ihr Ansprechpartner:

Benjamin Ledwon
Deutsche Telekom

E-Mail:
Benjamin.Ledwon@telekom.de

Ihr Ansprechpartner:

Taras Holoyad
Bundesnetzagentur

E-Mail:
Taras.Holoyad@BNetzA.DE