TÜV AI.LAB

Prüfung und Zertifizierung von KI in der Praxis: Herausforderungen und Lösungsansätze

Franziska Weindauer
Alexander von Janowski

Dialogveranstaltung | KI im Mittelstand
19.09.24

Agenda

1 Impuls: AI Act & Zertifizierung

2 Offene Diskussion zu KI & Zertifizierung

3 Übung: KI-Systeme klassifizieren

4 Technische Prüfung von Fairnessanforderungen

TÜV AI.LAB

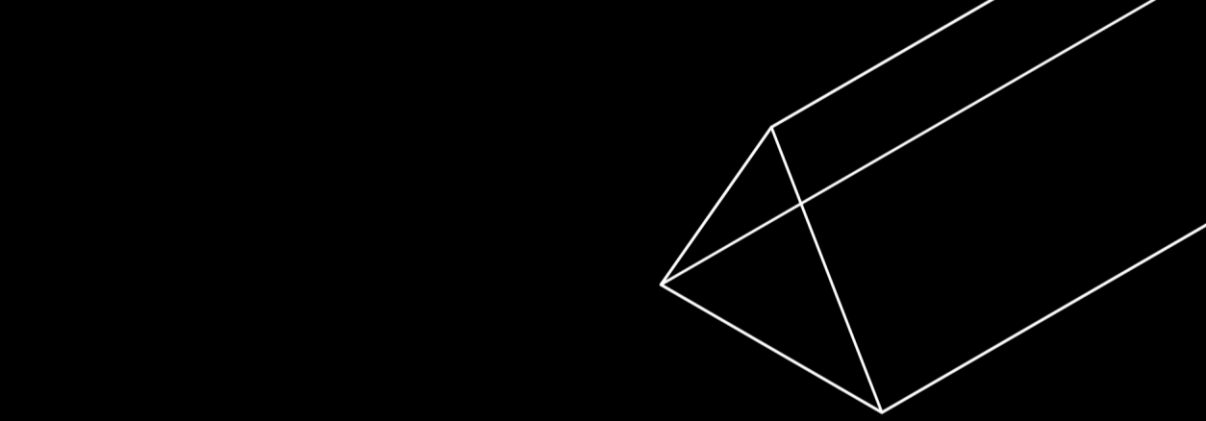
TÜV AI.LAB

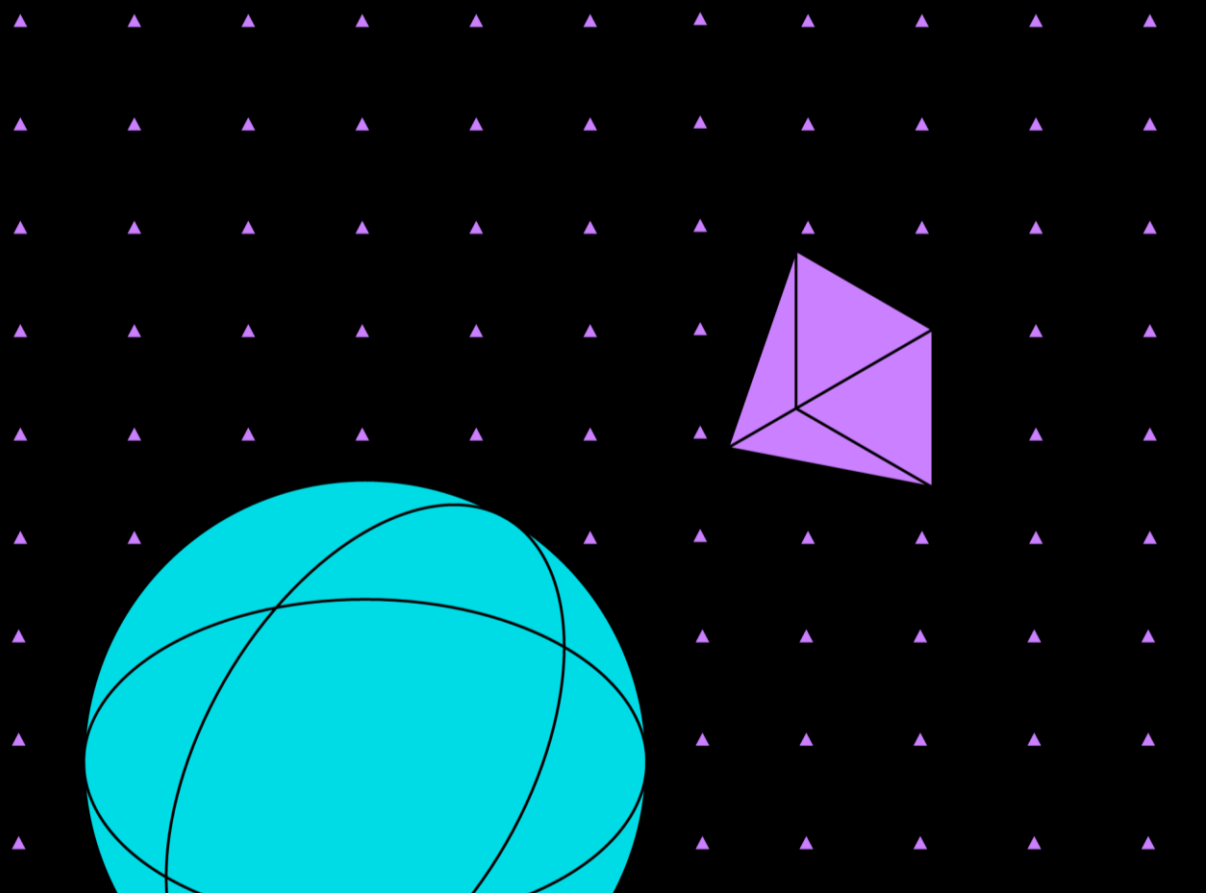
Teil I

-

EU AI Act

Überblick & Anforderungen

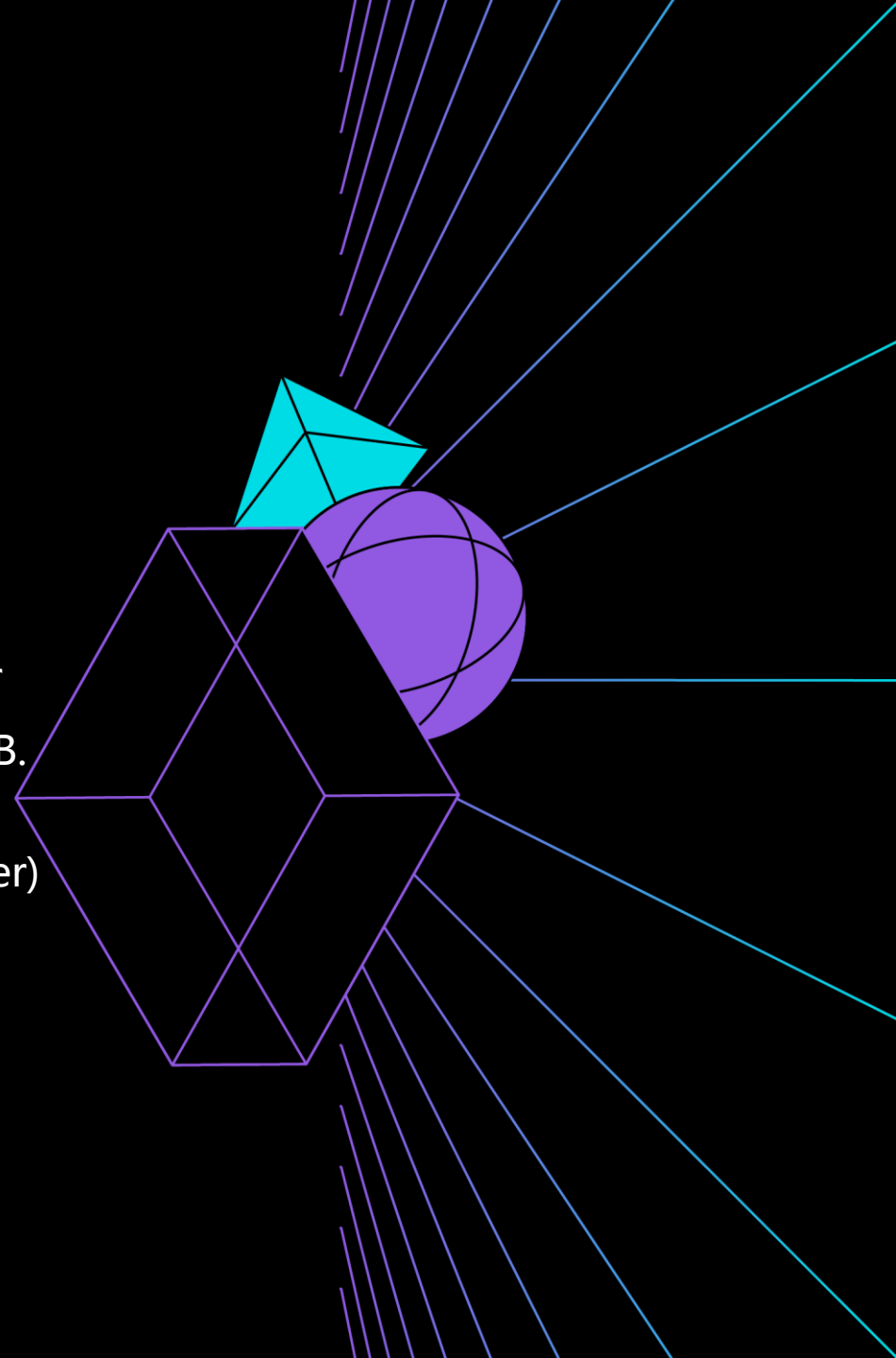




EU AI Act

Zentrale Fakten zur europäischen KI-Verordnung

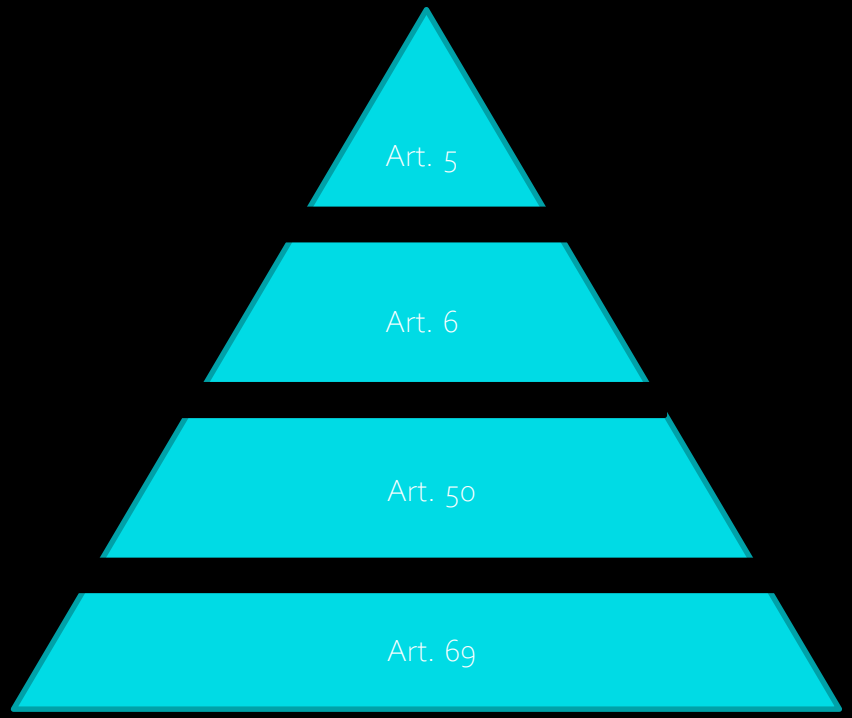
- Weltweit eine der ersten umfassenden Regulierungen für Künstliche Intelligenz
- Erstellt nach dem New Legislative Framework
- Risikobasierter Ansatz: Vier Risikoklassen, zzgl. General Purpose AI
- Regulierung von KI-Anwendungen, nicht der KI-Technologie als solcher
- Aussparung von Forschung/Entwicklung sowie bestimmter Bereiche (z.B. Militär)
- Definition verschiedener Akteure und Rollen (u.a. Anbieter und Betreiber)
- Zusammenwirken und Verschränkung mit anderen Verordnungen (z.B. DSGVO, Data Act) sowie mit sektoralen Regulierungen (z.B. Medical Device Regulation)



TÜV
AI.LAB

EU AI Act

Überblick über die Risikoklassen des EU AI Acts



Verbotene Systeme | Inakzeptables Risiko
Inverkehrbringung verboten (z.B. „Social Scoring“)

Hochrisiko-Systeme
Hohe Auflagen (z.B. für HR-Systeme)

Transparenzverpflichtungen | Begrenztes Risiko
Kennzeichnungspflichten (z.B. bei Chat-Bots)

Niedriges Risiko | Minimales Risiko
Keine Auflagen (z.B. Assistenz-KI zur Textoptimierung)

GPAI und Foundation Models

mit systemischem Risiko
u.a. Überwachungspflichten

ohne systemisches Risiko
v.a. Dokumentationspflichten

TÜV AI.LAB

Was sind Hochrisiko-Systeme?

Definition: Systeme, die ein hohes Risiko in Bezug auf Gesundheit, Sicherheit oder fundamentale Grundrechte von Personen darstellen, gemessen an der Schwere des möglichen Schadens und der Wahrscheinlichkeit seines Eintretens UND in von dieser Regulierung definierten Bereichen eingesetzt werden

Artikel 6: Hochrisiko Klassifizierung

KI-Systeme aus Anhang I, Abschnitt A, die als Sicherheitskomponente oder als Produkt verwendet werden sollen, und einer Konformitätsbewertung durch Dritte unterliegen

KI-Systeme aus Anhang III, die als Sicherheitskomponenten oder als eigenständiges Produkt verwendet werden sollen

Anhang I, Abschnitt A

- Maschinen
- Medizin-produkte
- Spielzeug

+ Unabhängige Drittprüfung

Anhang III (u.a.)

- Bio-metrie
- Kritische Infra-struktur
- Bildung
- Öffent-licher Dienst
- Kredit-vergabe
- Strafver-folgung

Ausnahmen zu Hochrisiko-Systemen

Ein KI-System in Anhang III wird nicht als Hochrisiko-System eingestuft, wenn eine der folgenden Bedingungen erfüllt ist:

1. Das KI-System ist dazu bestimmt, eine eng gefasste Verfahrensaufgabe durchzuführen;
2. Das KI-System ist dazu bestimmt, das Ergebnis einer zuvor abgeschlossenen menschlichen Tätigkeit zu verbessern;
3. Das KI-System ist dazu bestimmt, Entscheidungsmuster oder Abweichungen von früheren Entscheidungsmustern zu erkennen, und ist nicht dazu gedacht, die zuvor abgeschlossene menschliche Bewertung ohne eine angemessene menschliche Überprüfung zu ersetzen oder zu beeinflussen; oder
4. Das KI-System ist dazu bestimmt, eine vorbereitende Aufgabe für eine Bewertung durchzuführen, die für die Zwecke der in Anhang III aufgeführten Anwendungsfälle relevant ist

TÜV AI.LAB

Ausblick AI Act

TÜV AI.LAB

Der künftige Pfad des EU AI Act

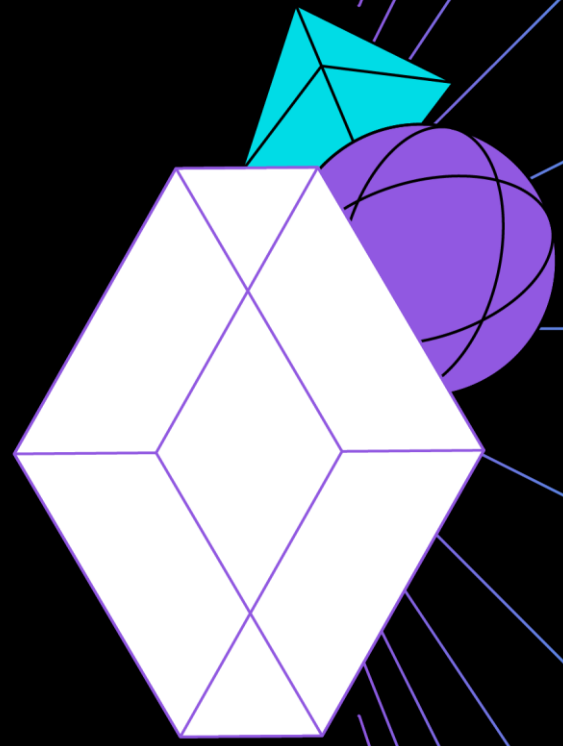
2024
2 Aug
Inkrafttreten AI Act

2025
2 Feb
Verbot von Systemen mit inakzetablem Risiko

2025
2 Aug
GPAI Anforderungen treten in Kraft

2026
2 Aug
Anforderungen für Annex 3

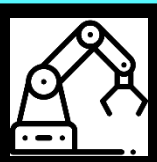
2027
2 Aug
Anforderungen für Annex 1A



AI.LAB

Vom AI Act zum Zertifikat

KI-Systeme, die in Bereiche mit unabhängiger Drittprüfung fallen



Sektorale Anforderungen (bspw. MDR)

⊕

AI Act Anforderungen

Unabhängige Drittprüfung

CE

KI-Systeme die in Hochrisiko-Bereiche nach Annex III fallen

⊗

Strafverfolgung, Migration, Biometrische Systeme

KI-Systeme die in Hochrisiko-Bereiche nach Annex III fallen

evtl.

Unabhängige Drittprüfung

Hersteller-intern

Trustworthy AI: Herausforderungen der KI-Prüfung

Für Unternehmen und Prüforganisationen

Schnelle Iterationszyklen bei KI-Systemen

- Schnelle Iterationszyklen und häufige, ggf. signifikante Änderungen machen kontinuierliche Überwachung und in Teilen Re-Zertifizierung notwendig
- KI-Systeme können sich ggf. selbst nach dem Deployment verändern / "weiterlernen"
- Zertifizierungsprozesse müssen überdacht werden.

Fehlende Ausfall-wahrscheinlichkeiten & Erfahrungswerte

- Bei KI-Modellen können keine Ausfallwahrscheinlichkeiten berechnet werden
- Es gibt wenige Statistiken und skalierbare Erfahrungswerte in der KI-Prüfung

Wenig standardisierte Testmethoden & Metriken

- Forschung im Feld der KI-Prüfung ist abhängig von der Prüfungsdimension und KI-Technologie zum Teil noch am Anfang (insb. GPAI)
- In vielen Bereichen keine bereits länger erprobten und gemeinhin bekannten Testverfahren sowie Metriken & Benchmarks

Dynamik in Regulierung, Technologie, Standards

- Regulierungslandschaft und Standardisierung dynamisch
- Schnelle technologische Fortschritte
- Dadurch kontinuierliche Anpassungsprozesse bei Entwicklern, Prüfern, Anwendern und anderen Akteuren erforderlich

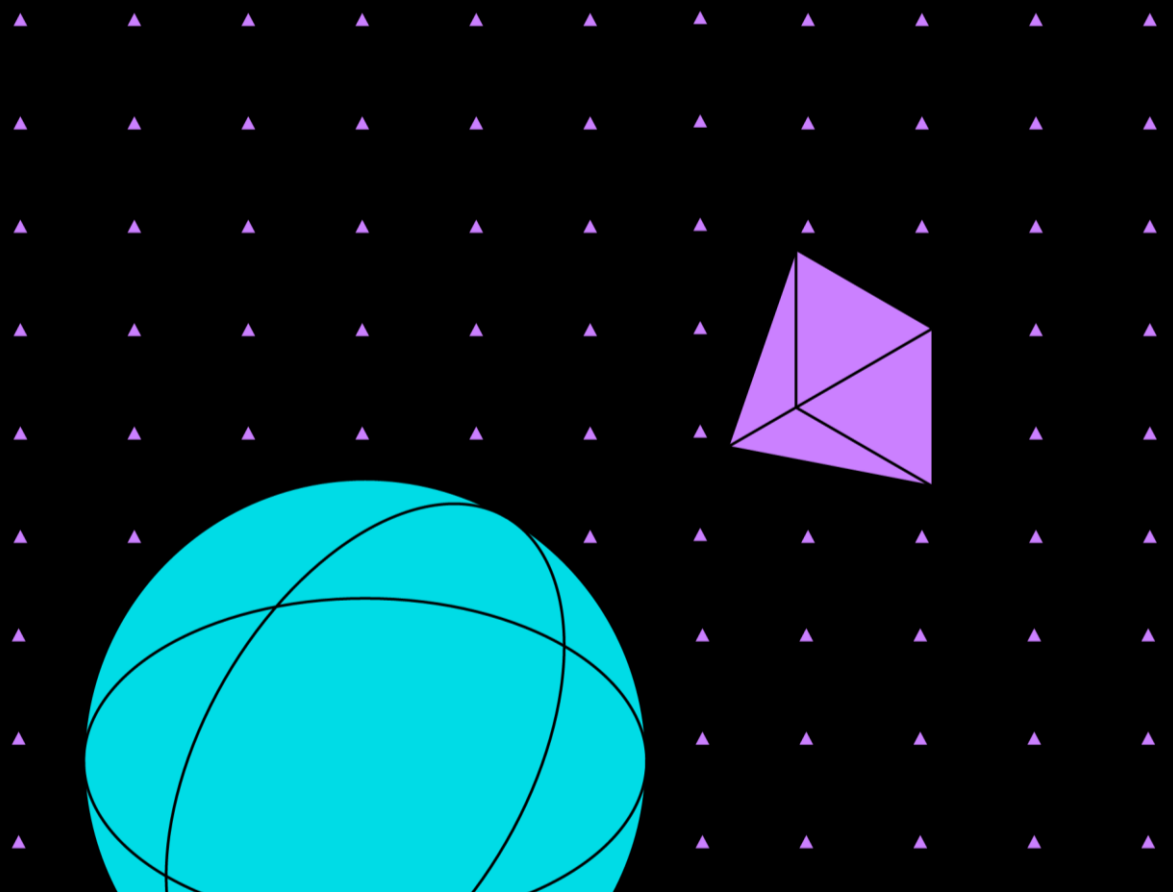
TÜV AI.LAB

TÜV AI.LAB

Teil II

-

Offene Diskussion

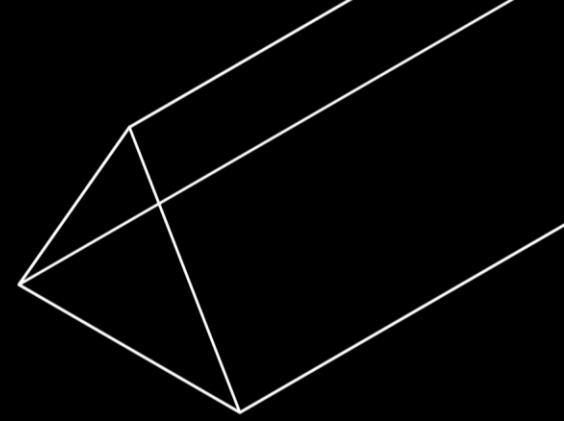


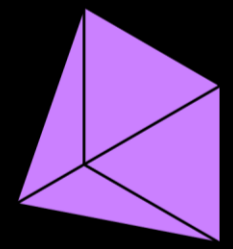
Vorstellungsrunde

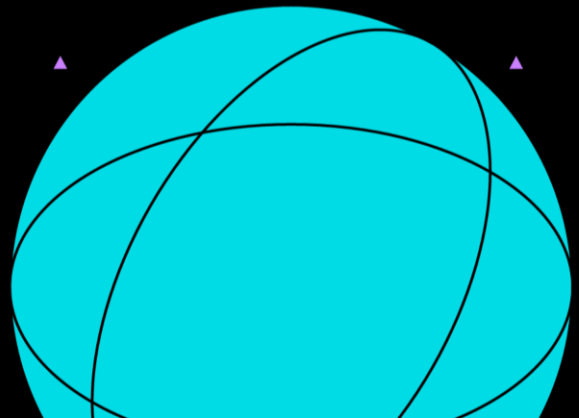
1 Wer sind Sie?

2 Was ist Ihr Bezug zu KI?

3 Was fasziniert Sie an KI am meisten aktuell?







TÜV AI.LAB

TÜV
AI.LAB

TÜV AI.LAB

Teil III

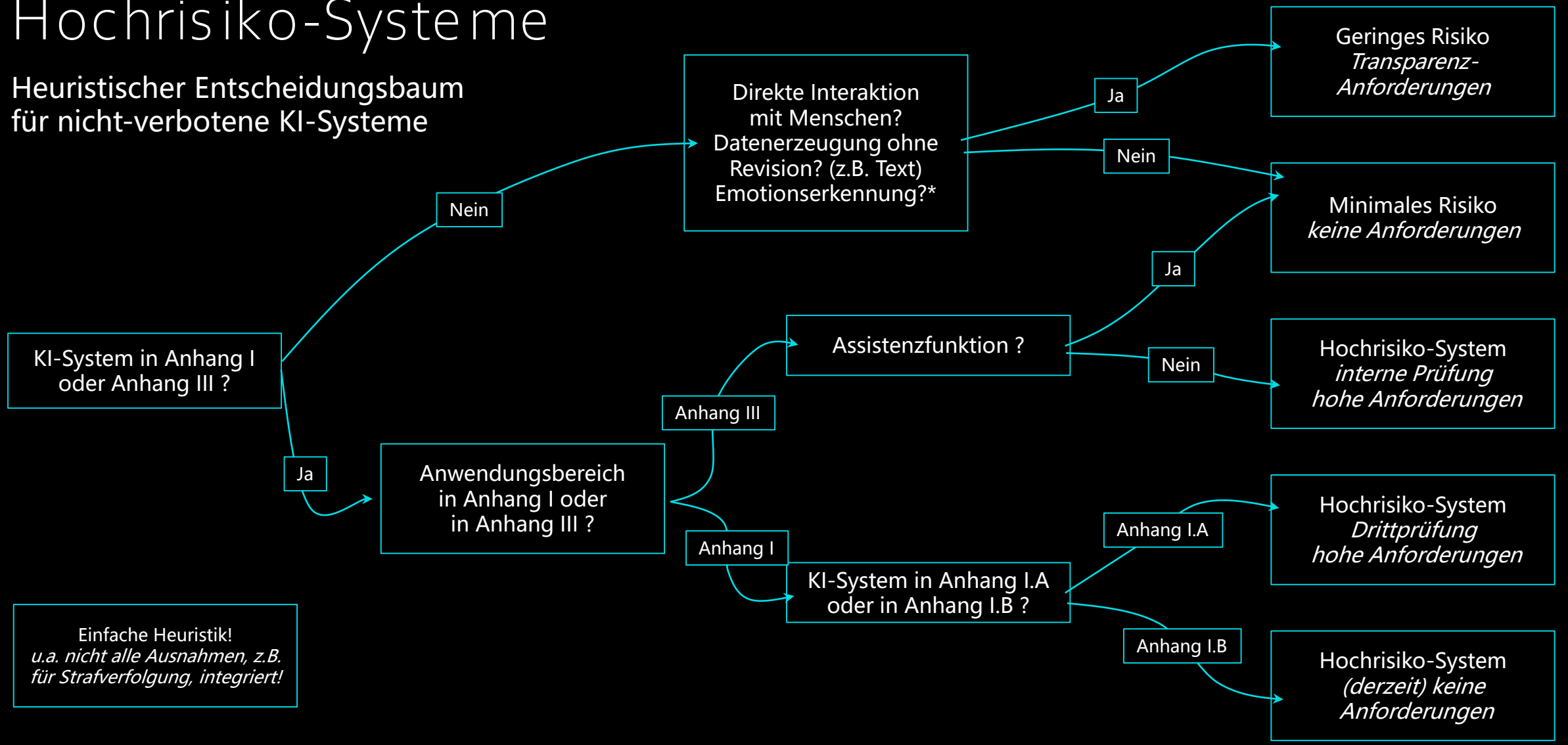
-

Übung zur AI Act Risikoklassifizierung

Dialogveranstaltung | KI im Mittelstand
19.09.24

Hochrisiko-Systeme

Heuristischer Entscheidungsbaum für nicht-verbotene KI-Systeme



**ausgenommen Emotionserkennung am Arbeitsplatz
**sowie in Bildungseinrichtungen (-> Verbotene Systeme)

Annex I, Abschnitt A

(Liste harmonisierter EU-Rechtsvorschriften)

1. Richtlinie 2006/42/EG – Maschinen
2. Richtlinie 2009/48/EG – Spielzeug
3. Richtlinie 2013/53/EU – Sportboote und Wassermotorräder
4. Richtlinie 2014/33/EU – Aufzüge und Sicherheitsbauteile für Aufzüge
5. Richtlinie 2014/34/EU – Geräte und Schutzsysteme in explosionsgefährdeten Bereichen
6. Richtlinie 2014/53/EU – Funkanlagen
7. Richtlinie 2014/68/EU – Druckgeräte
8. Verordnung (EU) 2016/424 – Seilbahnen
9. Verordnung (EU) 2016/425 – Persönliche Schutzausrüstungen
10. Verordnung (EU) 2016/426 – Geräte zur Verbrennung gasförmiger Brennstoffe
11. Verordnung (EU) 2017/745 – Medizinprodukte
12. Verordnung (EU) 2017/746 – In-vitro-Diagnostika

TÜV
AI.LAB

Annex I, Abschnitt B

(Liste anderer harmonisierter EU-Rechtsvorschriften)

1. Verordnung (EG) Nr. 300/2008 – Sicherheit in der Zivilluftfahrt
2. Verordnung (EU) Nr. 168/2013 – Zwei-, Dreirad- und Vierradfahrzeuge
3. Verordnung (EU) Nr. 167/2013 – Land- und forstwirtschaftliche Fahrzeuge
4. Richtlinie 2014/90/EU – Schiffsausrüstung
5. Richtlinie (EU) 2016/797 – Interoperabilität des Eisenbahnsystems
6. Verordnung (EU) 2018/858 – Kraftfahrzeuge und Kraftfahrzeuganhänger
7. Verordnung (EU) 2019/2144 – Typgenehmigung von Kraftfahrzeugen und deren Sicherheit
8. Verordnung (EU) 2018/1139 – Zivilluftfahrt und unbemannte Luftfahrzeuge

TÜV AI.LAB

Annex III

(Hochrisiko-KI-Systeme gemäß Artikel 6(2))

1. Biometrie (Fernidentifizierung, Kategorisierung, Emotionserkennung)
2. Kritische Infrastruktur
3. Bildung und Ausbildung (insbesondere Zulassungsentscheidungen, Bewertung/Benotung)
4. Zugang zu und Nutzung von privaten oder öffentlichen Dienstleistungen und Leistungen (z.B. Anspruch auf staatliche Unterstützung, Kreditwürdigkeit, Risikobewertung)
5. Strafverfolgung
6. Migration, Asyl und Grenzkontrolle
7. Justizverwaltung und demokratische Prozesse (Unterstützung der Justizbehörden, Beeinflussung des Wahlverhaltens/Wahlen)

TÜV AI.LAB

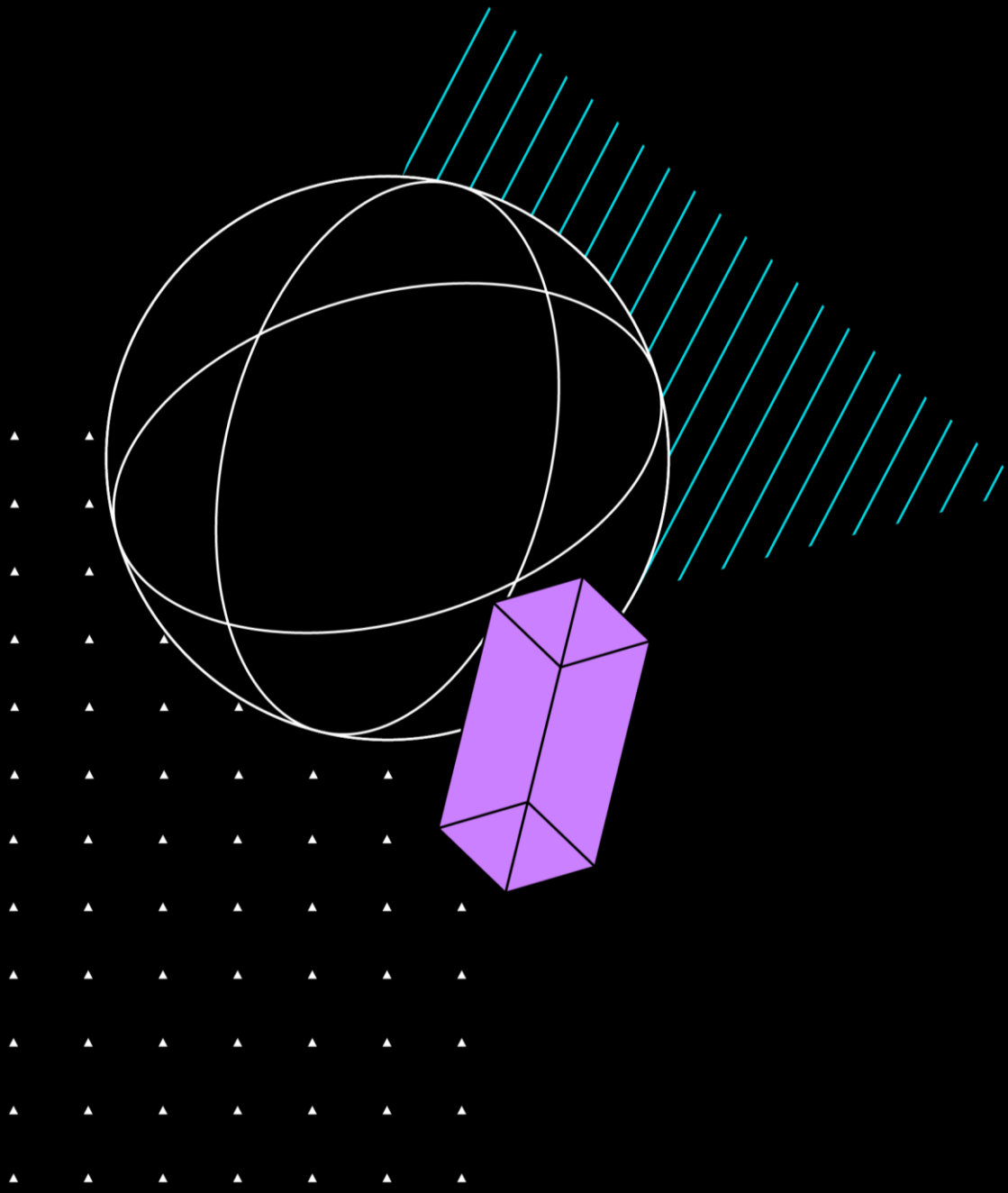
Ausnahmen von Annex III

1. Ein KI-System in Anhang III wird nicht als Hochrisiko-System eingestuft, wenn eine der folgenden Bedingungen erfüllt ist:

2. Das KI-System ist dazu bestimmt, eine eng gefasste Verfahrensaufgabe durchzuführen;
3. Das KI-System ist dazu bestimmt, das Ergebnis einer zuvor abgeschlossenen menschlichen Tätigkeit zu verbessern;
4. Das KI-System ist dazu bestimmt, Entscheidungsmuster oder Abweichungen von früheren Entscheidungsmustern zu erkennen, und ist nicht dazu gedacht, die zuvor abgeschlossene menschliche Bewertung ohne eine angemessene menschliche Überprüfung zu ersetzen oder zu beeinflussen; oder
5. Das KI-System ist dazu bestimmt, eine vorbereitende Aufgabe für eine Bewertung durchzuführen, die für die Zwecke der in Anhang III aufgeführten Anwendungsfälle relevant ist

TÜV AI.LAB

Vielen Dank



Alexander von Janowski
Manager AI Certification
alexander@tuev-lab.ai

TÜV AI.LAB