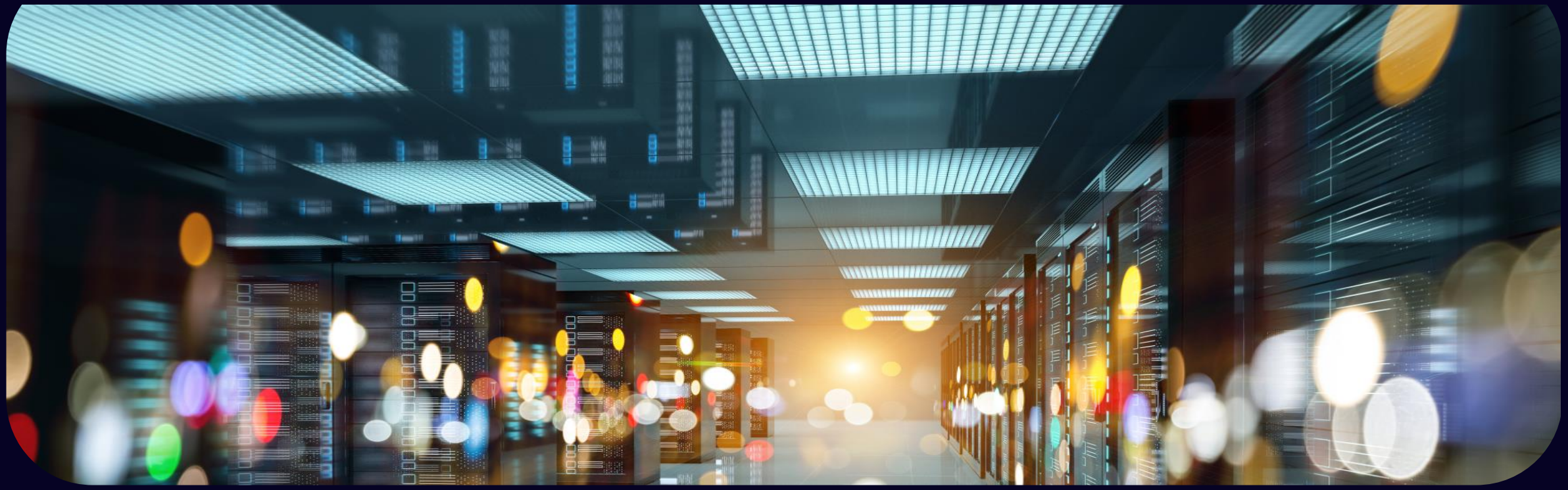


PROCON IT

EU AI Act

Christoph Hoffmann, April 2024

Über den Vortragenden und den Vortrag

DISCLAIMER

Christoph Hoffmann, Head of Data Solutions & AI, PROCON IT

Studium: Statistik, VWL, Philosophie, NICHT JURA!

Kontakt: christoph.hoffmann@procon-it.de

Keine Rechtsberatung!
Diskretionäre Lücken in der Darstellung!
Unsicherheit in der Auslegung!

Erfahrung in Umsetzungsunterstützung
Vermittlung der Kerninhalte
Startpunkt für Diskussionen & Recherche

Agenda

Bist du betroffen?

Was kommt auf dich zu?

Was kannst du tun?

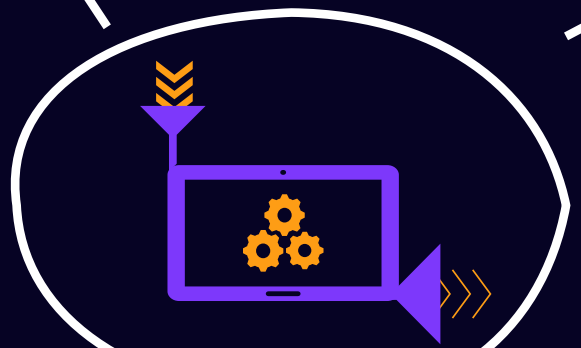
Warum und wer?

Warum?

- Funktionieren des Binnenmarktes durch einheitlichen Rechtsrahmen verbessern (Fragmentierung verhindern)
- Schutz von natürlichen Personen, Unternehmen, Demokratie, Rechtsstaatlichkeit, Umwelt
- Einführung von menschenzentrierter und vertrauenswürdiger KI fördern (Einheitliches Schutzniveau)

Für wen?	Anbieter in der Union	Betreiber in der Union	Anbieter & Betreiber weltweit wenn Output in Union	Einführer, Bevollmächtigte, Produkthersteller, …
Für wen nicht?	Militär, Verteidigung, Nationale Sicherheit	Wissenschaftliche Forschung	Privatpersonen als Betreiber	Quelloffen, wenn nicht Hochrisiko-KI

Was ist ein KI-System (Artikel 3 + EW 12)?



maschinengestütztes System

Implizite oder explizite Ziele

in unterschiedlichem Grade autonomen Betrieb

nicht Systeme die menschengemachte Regeln automatisiert ausführen (frei nach EW 12)

Standalone product or component (EW 12)

key characteristic of AI systems ... is to **infer** outputs, such as predictions ... or decisions... from inputs or data (EW 12)

physische oder virtuelle Umgebungen beeinflussen können

Techniques include machine learning, logic- and knowledge based approaches (EW 12)

Kann anpassungsfähig sein (self-learning, EW 12)

transcends basic data processing by enabling learning, reasoning or Modelling (EW 12)

KI-Systeme – 3 Beispiele

Autohersteller trainiert ein ML-Modell um Schadensfälle durch Sensor-Input vorherzusagen. Die Software zeigt dem Autofahrer eine Wartungsindikation an.

Obiges ML-Modell sendet Wartungs-Prognose an Service-Mitarbeiter welcher den Fall prüft und festgestellte Wartungsindikation an Autofahrer schickt.

Softwareentwickler schreibt Software die die Wartungsindikation anzeigt, wenn Sensorausschlag größer X ist.

„Ist das KI oder kann das weg?“

Sind wir ein Betreiber (Deployer)?

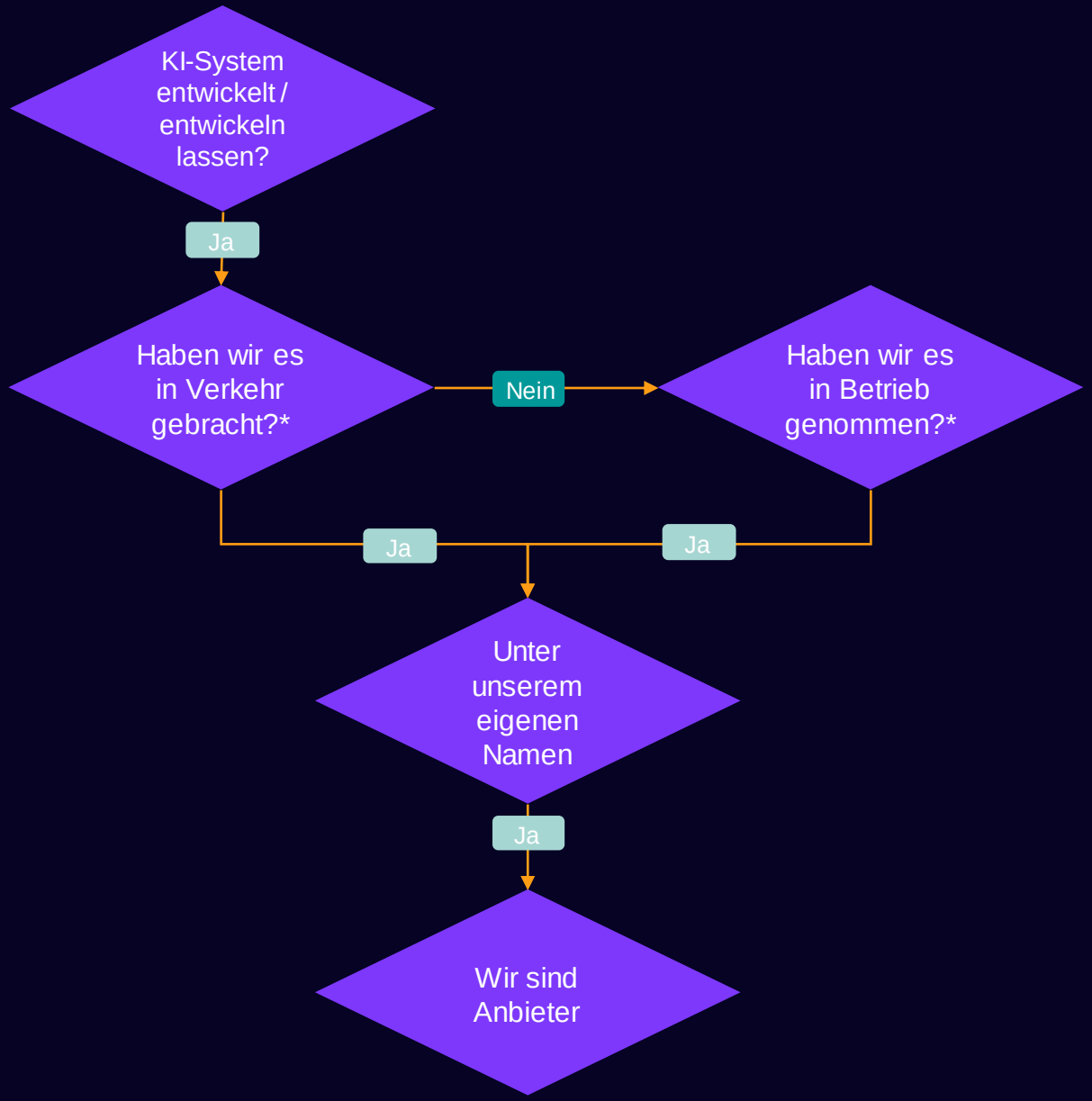
Art.3 Nr. 4

Eine natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System in eigener Verantwortung („under its authority “) **verwendet**, es sei denn, das KI-System wird im Rahmen einer persönlichen und nicht beruflichen Tätigkeit verwendet;

Ab wann gilt man unter der KI-Verordnung als Betreiber? Der Einsatz eines KI-Systems im Unternehmen, z.B. sogar die Verwendung von ChatGPT, gilt nach m.E. als „betreiben“ (Sonst keine Transparenzpflicht).

KI-gestützte Kreditrisikoprüfsoftware in einer Bank-Filiale offensichtlich auch.

Sind wir ein KI-Anbieter (Provider)?



Inverkehrbringen

Erstmalige Bereitstellung eines KI-Systems auf dem EU Markt

Inbetriebnahme

Bereitstellung direkt an Betreiber oder zum Eigengebrauch

KI Anbieter und Betreiber – 3 Beispiele

Autohersteller entwickelt KI System um vorherzusagen ob ein Motor einer Wartung bedarf. Das KI-Modul läuft im Auto.

Anbieter?
Betreiber?
Endnutzer?

Software-Entwicklungsfirma entwickelt KI System um vorherzusagen ob ein Motor einer Wartung bedarf. Software wird an Autohersteller verkauft.

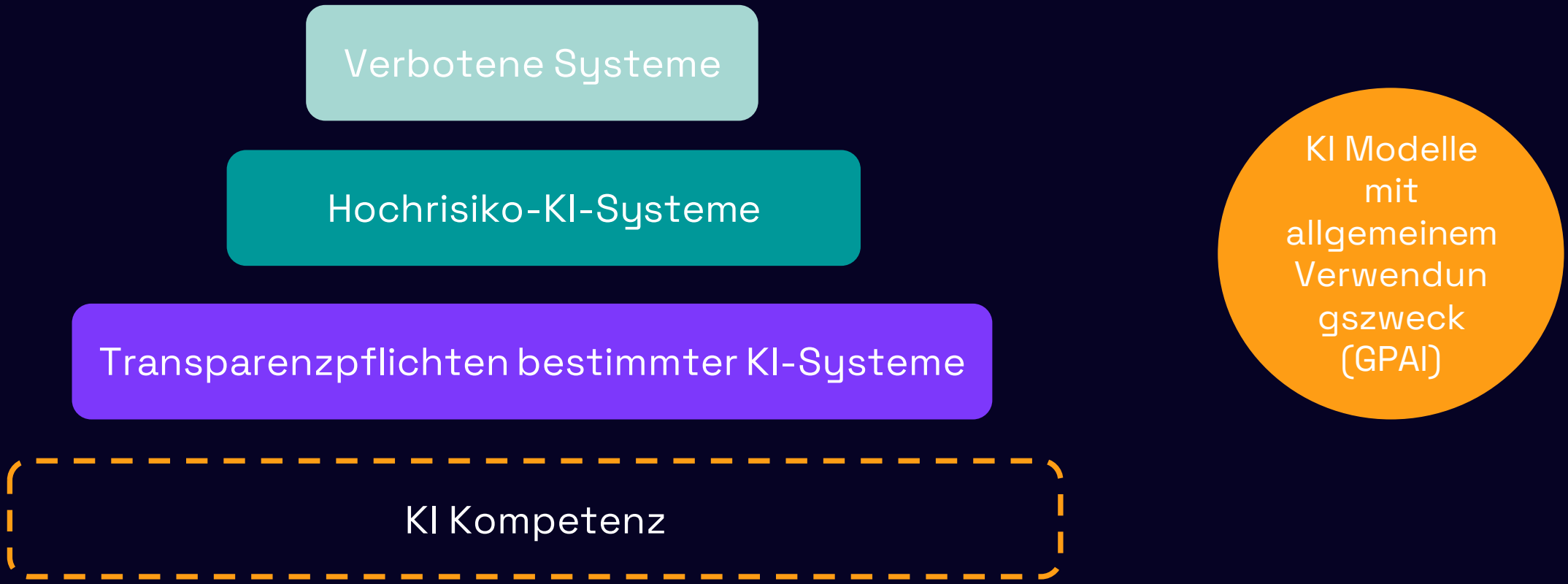
Softwarefirma?
Autohersteller?

Was wenn AH Entwicklung beauftragt?

KI-Entwickler einer Beratung hilft Autohersteller bei Entwicklung eines KI System das vorherzusagt ob ein Motor einer Wartung bedarf.

Softwarefirma?
KI-Entwickler?
Autohersteller?

Risikobasierter Regulierungs-Ansatz



KI Kompetenz (AI literacy, Art.4)

Wer

Alle Anbieter und alle Betreiber

Was

- Maßnahmen um sicherzustellen das involviertes Personal ausreichende KI-Kompetenz hat
- Verständnis für Rechte & Pflichten der Regulierung (Art. 3, Nr. 56)
- Bewusstsein für Chancen & Risiken (Art. 3, Nr. 56)

Wie (Eigene Ideen)

- KI-Richtlinie im Unternehmen
- Regelmäßiges Schulungen KI für ausgewählte Mitarbeiter:
  - Modul - KI Basics
  - Modul – Vertrauensvolle KI
  - Modul – KI Verordnung
  - Modul – Spezifisches KI-System
- Dokumentation der Schulungsteilnehmer

KI-Modelle mit allgemeinem Verwendungszweck (GPAI)

- Erhebliche allgemeine Verwendbarkeit
- Breites Spektrum unterschiedlicher Aufgaben
- Integration in Vielzahl nachgelagerter Systeme oder Anwendungen möglich
- Ausgenommen Modelle vor Inverkehrbringen für R&D

- Hohe Reichweite
- Negative Folgen für Gesundheit, Sicherheit & Gesellschaft
- Effekte über gesamte Wertschöpfungskette

GPAI mit systemischen Risiko

Anforderungen für Anbieter von GPAIS

Technische Dokumentation (Trainings, Testverfahren,...)

Anbieter von KI-Systemen informieren

Urheberrechtsstrategie

Veröffentlichung zu Trainingsdaten

Was für GPAIs gilt +

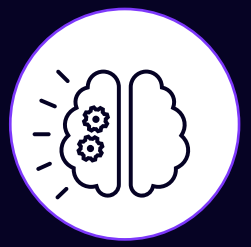
Anzeigen nach spätestens 2 Wochen

Modellbewertung (Angriffstests)

Risikobewertung & -Minderung

Cybersicherheit

Verbotene Praktiken im KI-Bereich (Prohibited AI Practices, Art.5)



Manipulation mit erheblichen Schaden

Schutzbedürftigkeit ausnutzen

Social Scoring

Straftatvorhersage durch Profiling

Wahllos Gesichts-erkennung

Emotionen am Arbeitsplatz / Bildung

Biometrische Kategorisierung

Strafverfolgung öffentlicher Raum (Real-time)

Auswahl an Hochrisiko-KI-Systemen (Art.6)

Produkte / Sicherheitsbauteile nach Anhang I

Spielzeug, Sportboote, Aufzüge, Funkanlagen, Seilbahnen, Medizinprodukte, Zivilluftfahrt, Fahrzeuge (zwei, drei, vierrädrig),...

Ausnahmen Anhang III

Abweichungen erkennen

Enge Verfahrensaufgabe

Abgeschlossene Aufgabe verbessern,

Vorbereitende Aufgabe für Bewertung

Bereiche nach Anhang II

- Biometrie (Fernidentifizierung, Kategorisierung, Emotionserkennung)
- Sicherheitsbauteile in kritische Infrastruktur (digital, Straßenverkehr, Wasser-, Gas-, Stromversorgung)
- Bildung (Zulassung, Bewertung Lernergebnisse, Prüfungsüberwachung, ...)
- Beschäftigung / Personal (Einstellung, Bewerberauswahl, Beförderungen, Kündigungen, Aufgabenzuweisung)
- Inanspruchnahme grundlegender öffentlicher/privater Dienste (Zugang Unterstützungsleistungen, Gesundheitsdienste, Kreditwürdigkeit, Preisbildung/Risikobewertung Lebens- und Krankenversicherung)
- Strafverfolgung
- Migration, Asyl, Grenzkontrollen
- Rechtspflege / Demokratische Prozesse

Anforderungen Hochrisiko-KI-Systeme (Auswahl)

Anbieter

- Risikomanagementsysteme
- Daten & Datengovernance
- Technische Dokumentation
- Aufzeichnungspflichten
- Informationen für Betreiber
- Menschliche Aufsicht
- Genauigkeit, Robustheit, Cybersicherheit
- Qualitätsmanagementsystem
- Verantwortung KI Wertschöpfungskette
- Erfassung in Datenbank

Betreiber

- Ordnungsgemäßer Betrieb
- Input-Prüfung
- Menschliche Aufsicht
- Protokollierung

Transparenzpflichten bestimmter KI-Systeme

Anbieter:

- Betroffene Personen können per Design erkennen das sie mit KI interagieren, außer es ist offenkundig
- Synthetisch generierte Audio-, Bild-, Video- oder Textinhalte sind als solche gekennzeichnet

Betreiber:

- Natürliche Person wird über Emotionserkennungssystem informiert
- Offenlegungspflicht wenn Bild, - Ton- oder Videoinhalte, die ein Deepfake sind, das Inhalte künstlich erzeugt wurden
- Veröffentlichte generierte Texte von öffentlichem Interesse sind kennzeichnungspflichtig

Sanktionen

Falsche Informationen bis zu 7,5 Mio. oder 1% Umsatz

Verstoß gegen Verbote bis zu 35 Mio. oder 7% Umsatz

Verstoß gegen Bestimmung bis zu 15 Mio. oder 3% Umsatz

Bei Sanktionen Interessen von KMUs & Startups berücksichtigen

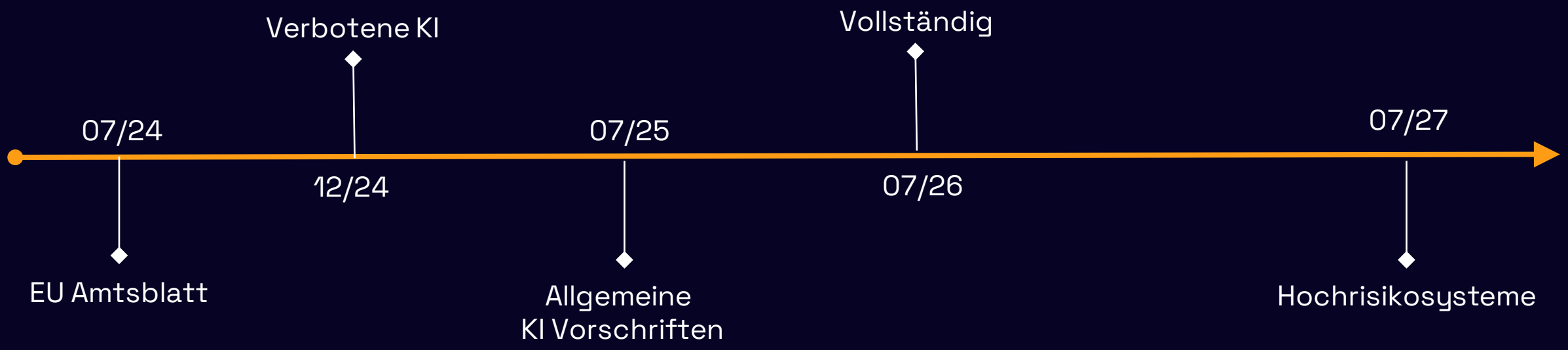
Ermessensspielraum der Behörde

Milderung durch Kooperation

Zeitrahmen

In Kraft treten wird Ende der aktuellen Legislaturperiode erwartet.

Danach Übergansfristen nach X Monaten, veranschaulicht mit Start Mitte 2024



Next Steps

Inventarisierung

Fiktives Beispiel

Model	Zweck	Owner	Abteilung	Verfahren		Selbstl
Credit-Scorer	Kreditwürdigkeitsprüfung	C. H.	Kreditabteilung	Random Forest	Mittel	
Absatz-Predictor	Prognose Verkaufszahlen	Peter H.	Einkauf	Neuromales Netz	Manuel	Nein
Customer-Chatbot	Produktnachfrage	Julia M.	Customer Service	Open AI ChatGPT	Hoch	Nein

Klassifizierung

Model	Klassifizierung des Risikos	Rolle
Credit-Scorer		Betreiber, Anbieter
Absatz-Predictor		Anbieter
Customer-Chatbot		Betreiber

Maßnahmen

Model	Klassifizierung des Risikos	Business Impact	Code of Conduct	Transparenz	Dokumentation
Credit-Scorer		Hoch	✓	✓	✗
Absatz-Predictor		Mittel			
Customer-Chatbot		Gering	-	✗	✗

Inventarisierung + Klassifizierung + Maßnahmenkatalog auf Kosten-Nutzen-Basis

Vielen Dank für eure Aufmerksamkeit!