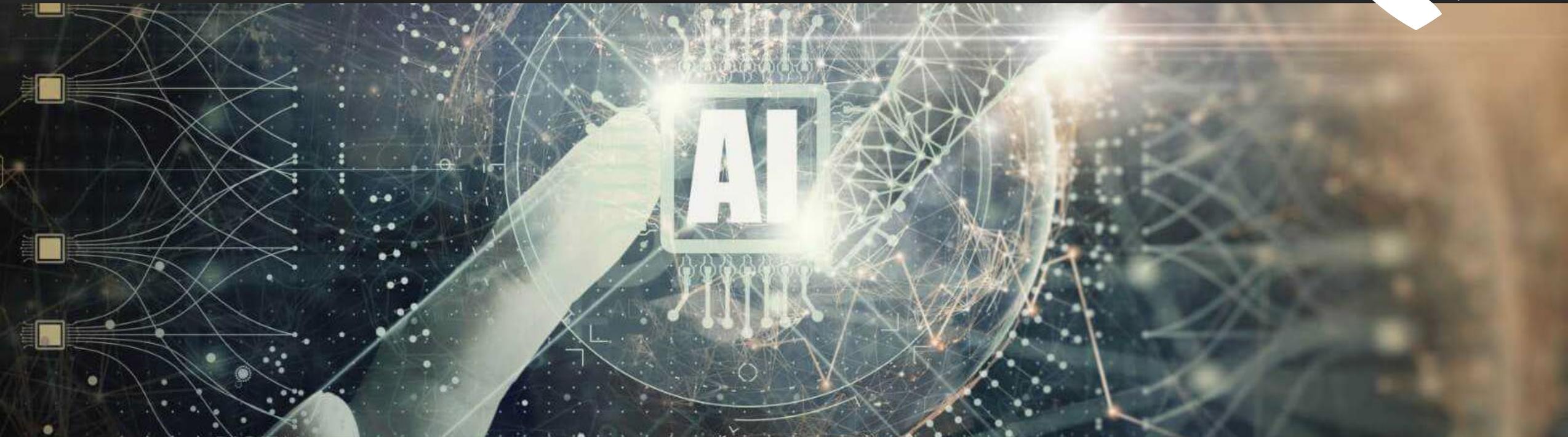


Künstliche Intelligenz verantwortungsvoll einsetzen

Ethische und regulative Herausforderungen



Karin Wegenstein | karin.wegenstein@fhwn.ac.at
Ralph Sichler | ralph.sichler@fhwn.ac.at

FACHHOCHSCHULE WIENER NEUSTADT

Business Tagung 2023

KI-Ethics



- ▶ Beispiele von KI-Anwendungen, die unsere Vorstellungen von Ethik und Moral herausfordern sind
 - ▶ KI Anwendung im Recruiting zur Selektion von Bewerbungen
 - ▶ KI Anwendungen, die Entlassungen vorschlagen
 - ▶ KI Anwendung zur Gesichtserkennung
 - ▶ Emotionserkennung, beispielsweise im Sales-Setting
 - ▶ Social Scoring
 - ▶ KI-Einsatz zu Überwachungszwecken, sei es im öffentlichen Raum oder am Arbeitsplatz

KI-Ethics



- ▶ Beispiele von KI-Anwendungen, die unsere Vorstellungen von Ethik und Moral herausfordern sind
 - ▶ KI-Anwendung zur Beeinflussung von Personen (inkl. öffentlicher Meinung, Manipulation demokratischer Prozesse)
 - ▶ KI-Anwendung mit potenziellen Personenschaden (Bsp. autonom fahrender Autos, sicherheitskritische Anwendungen)
 - ▶ KI-Anwendungen mit hohem Energieverbrauch
 - ▶ KI-Anwendungen zu Vorhersagezwecken, die ggf. die Stabilität z.B. von Märkten und Währungen beeinflussen können



KI schafft neue Realitäten und
wirft dadurch neue Fragen auf.



KI-Anwendung mag unsere gewohnten Vorstellungen von Ethik und Moral herausfordern

Verantwortung hinsichtlich der **Gestaltung** von KI

Verantwortung hinsichtlich der **Nutzung** von KI

Bedarf eines **Regulativs**

Werkzeuge um KI zu gestalten



Datengrundlage
gestalten



Feature
Engineering,
Feature Selection



Algorithmus
gestalten &
Dokumentation



Hyperparameter
Tuning



Fehlerreduktion



Testen, Output-
Analysen



Energie-Effizienz
sicherstellen



Sicherheit
gewährleisten



KI-Anwendung mag unsere gewohnten Vorstellungen von Ethik und Moral herausfordern

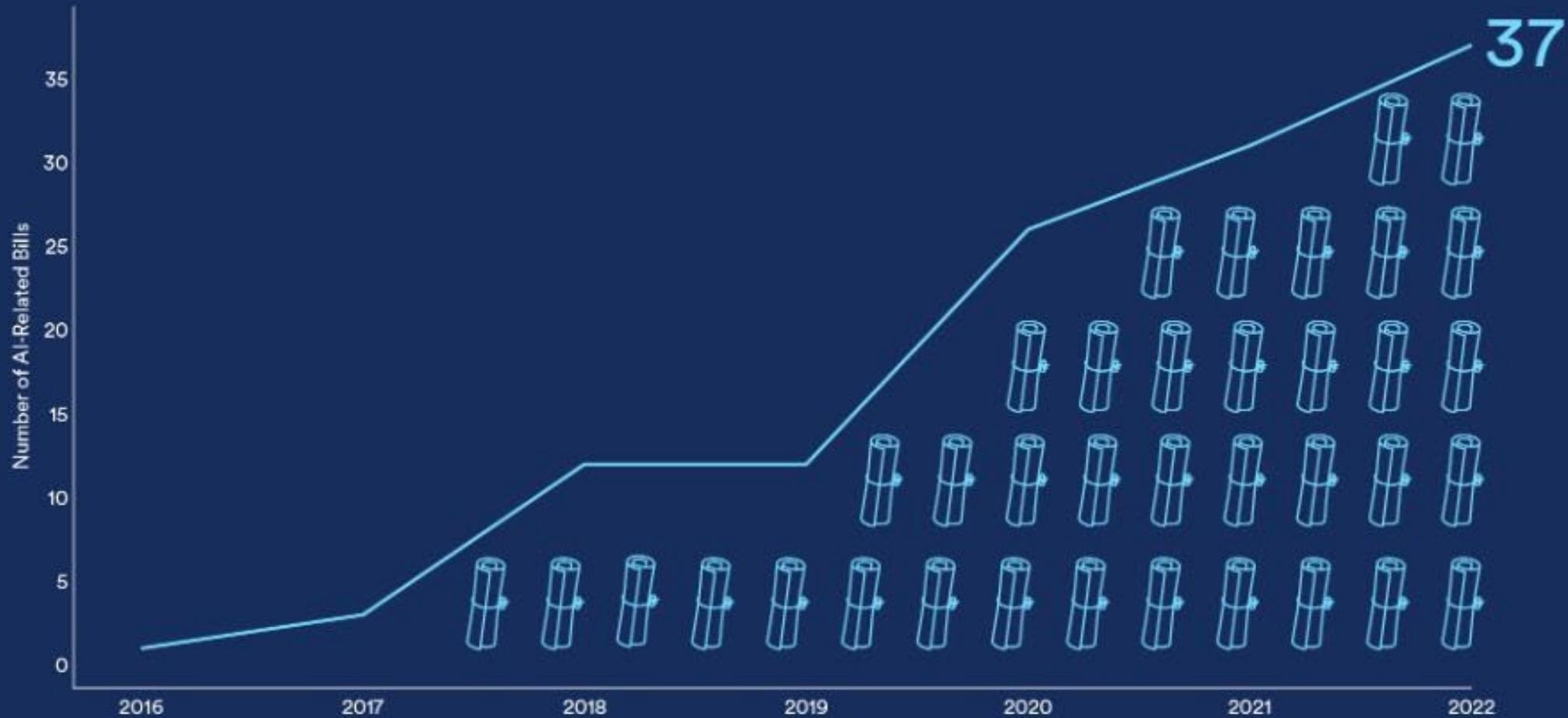
Verantwortung hinsichtlich der **Gestaltung** von KI

Verantwortung hinsichtlich der **Nutzung** von KI

Bedarf eines **Regulativs**



Number of AI-Related Bills Passed into Law Globally



Source: AI Index, 2022 | Chart: 2023 AI Index Report

*Image:
Stanford
University
2023 AI
Index Report*

USA: AI Bill of Rights



- ▶ Informativen Charakter, keine Rechtsverbindlichkeit
- ▶ Keine umfassende KI-Regulierung

- ▶ Entwürfe für umfassende Regulierung von KI gibt es aktuelle bspw. in Brasilien, Kanada und in der EU



EU Artificial Intelligence Act

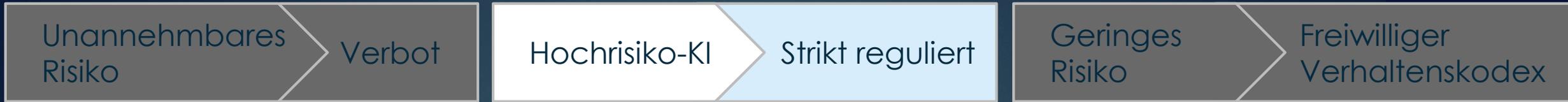
Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz

EU AI Act: Risikobasierter Ansatz



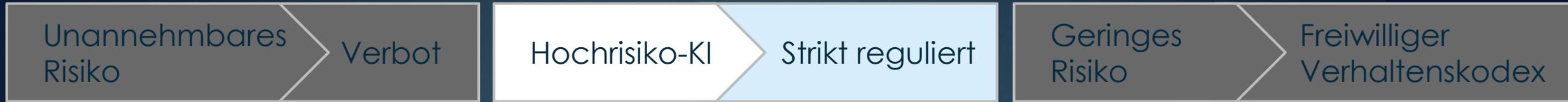
- ▶ Unbewusste Beeinflussung oder Beeinflussung schutzbedürftiger Personen mit möglicher Schadensfolge (psychisch oder physisch) für dieselbe oder eine andere Person
- ▶ Klassifizierung von Vertrauenswürdigkeit aufgrund des Sozialverhaltens durch Behörden, aus der eine Benachteiligung resultieren kann
- ▶ Echtzeit-Fernidentifikation, außer zum Abwenden besonders schwerer Verbrechen

EU AI Act: Risikobasierter Ansatz



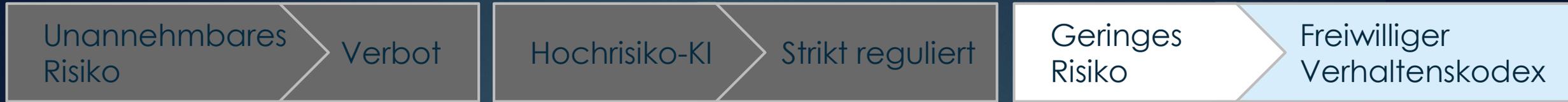
- ▶ Anwendungen, die Sicherheit von Produkten (Maschinen, Spielzeuge, Medizinprodukte) und Transportmittel (Autos, Züge, Flugzeuge, etc.) gewährleisten sollen
- ▶ Anwendungen in Zusammenhang mit kritischer Infrastruktur
- ▶ Biometrische ID-Systeme
- ▶ Systeme, die den Bildungszugang, Werdegang und die Anstellung einer Person beeinflussen können (z.B. Leistungsbeurteilung durch KI, Recruitment)
- ▶ Essentielle private und öffentliche Services (Kreditvergabe, Vergabe von Sozialleistungen)
- ▶ Gesetzesvollzug-Systeme, die in Konflikt mit Grundrechten kommen könnten (e.g. pre-crime detection)
- ▶ Migrationsmanagementsysteme (Verifikation von Dokumenten)
- ▶ Justiz-Anwendungen (Straf-Assistenzsysteme)

EU AI Act: Risikobasierter Ansatz



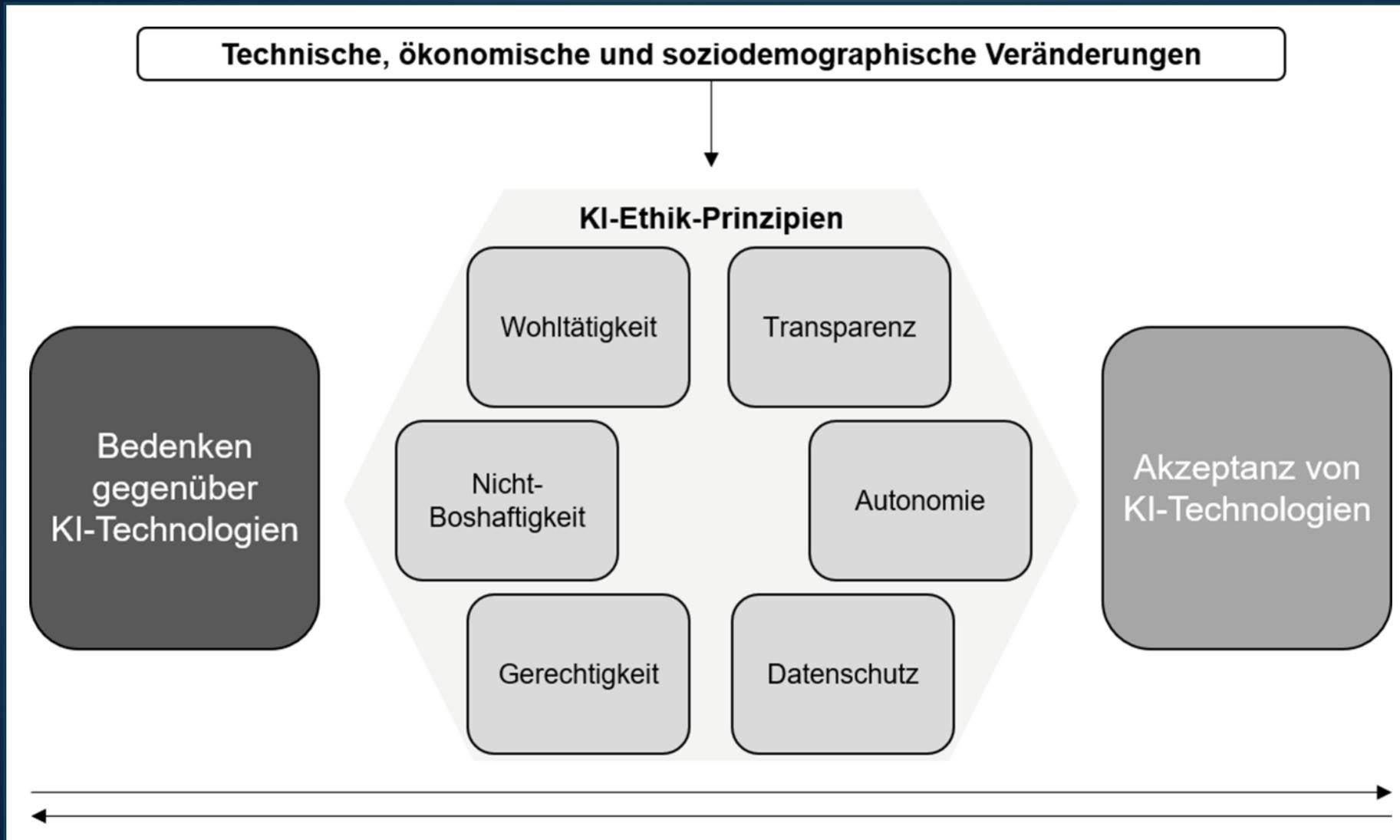
- ▶ Meldepflicht – EU-Datenbank für Hochrisiko-KI-Systeme
- ▶ Beobachtung/Überwachung der Systeme
- ▶ CE-Zertifizierung
- ▶ Spezifische Anforderungen hinsichtlich Verwendung von Daten und Data Governance, technische Dokumentation, Transparenz, Information für Anwender*innen, menschliche Überwachung, Robustheit, Genauigkeit und Sicherheit
- ▶ Risikomanagementsystem

EU AI Act: Risikobasierter Ansatz



- ▶ **Transparenzpflichten:**
 - ▶ Nutzer*innen sind informiert, dass sie mit KI interagieren
 - ▶ Inhalte sind entsprechend gekennzeichnet (Deepfake)
- ▶ Freiwillige Erfüllung der Anforderungen an Hochrisiko-KIs
- ▶ Verhaltenskodex

Prinzipien für eine KI Ethik



(Barton/
Pöppelbuß 2022)

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems



- ▶ **Sustainable development**
- ▶ Well-being
- ▶ Human-centered values
- ▶ Fairness
- ▶ Transparency and explainability
- ▶ **Robustness, security** and safety
- ▶ **Accountability**

High-Level Expert Group on AI of the European Commission



- ▶ Ethical Guidelines for Trustworthy AI
 - ▶ AI should be
 - ▶ Lawful
 - ▶ Ethical → **Human Agency & Oversight**
 - ▶ Robust
 - ▶ Technical Robustness & Safety
 - ▶ Privacy & Data Governance
 - ▶ Transparency
 - ▶ **Diversity**, Non-Discrimination & Fairness
 - ▶ **Societal & Environmental Well-Being**
 - ▶ Accountability

Grundkonzepte der Ethik

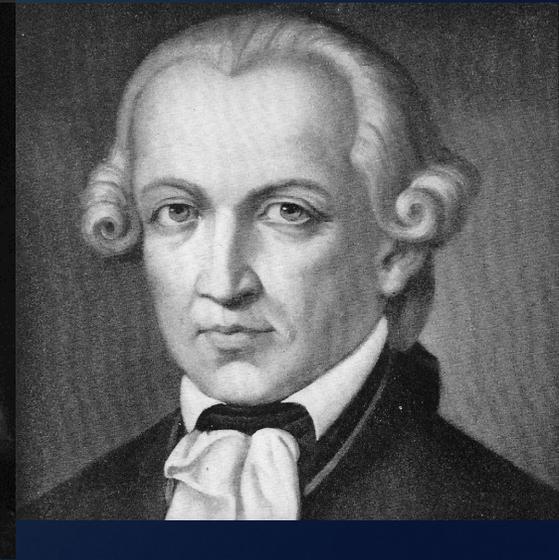
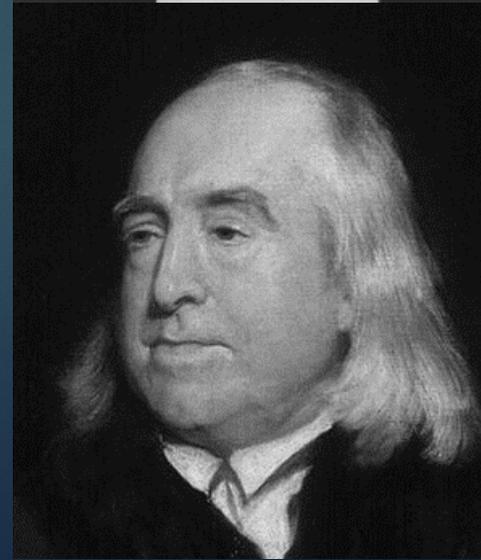
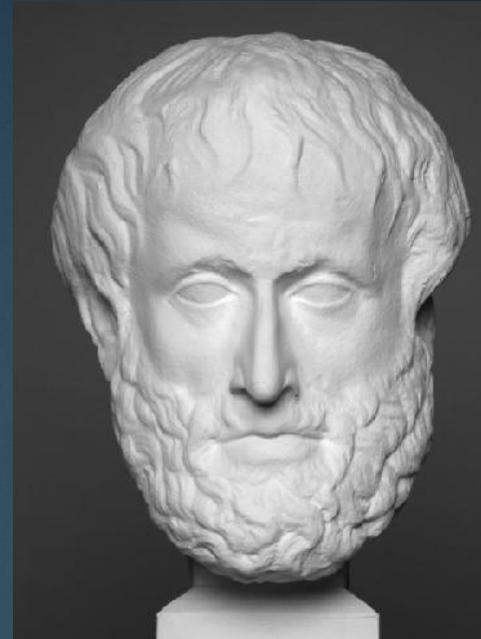


Die drei klassischen Grundpositionen der normativen Ethik

Tugendethik: Gut sein und gut leben!

Pflichtenethik: Gutes wollen und tun!

Konsequentialismus: Gutes bewirken!



Was bedeutet dies für die Digitale Ethik?



Leben

Kann KI zu einem besseren **Leben** führen?

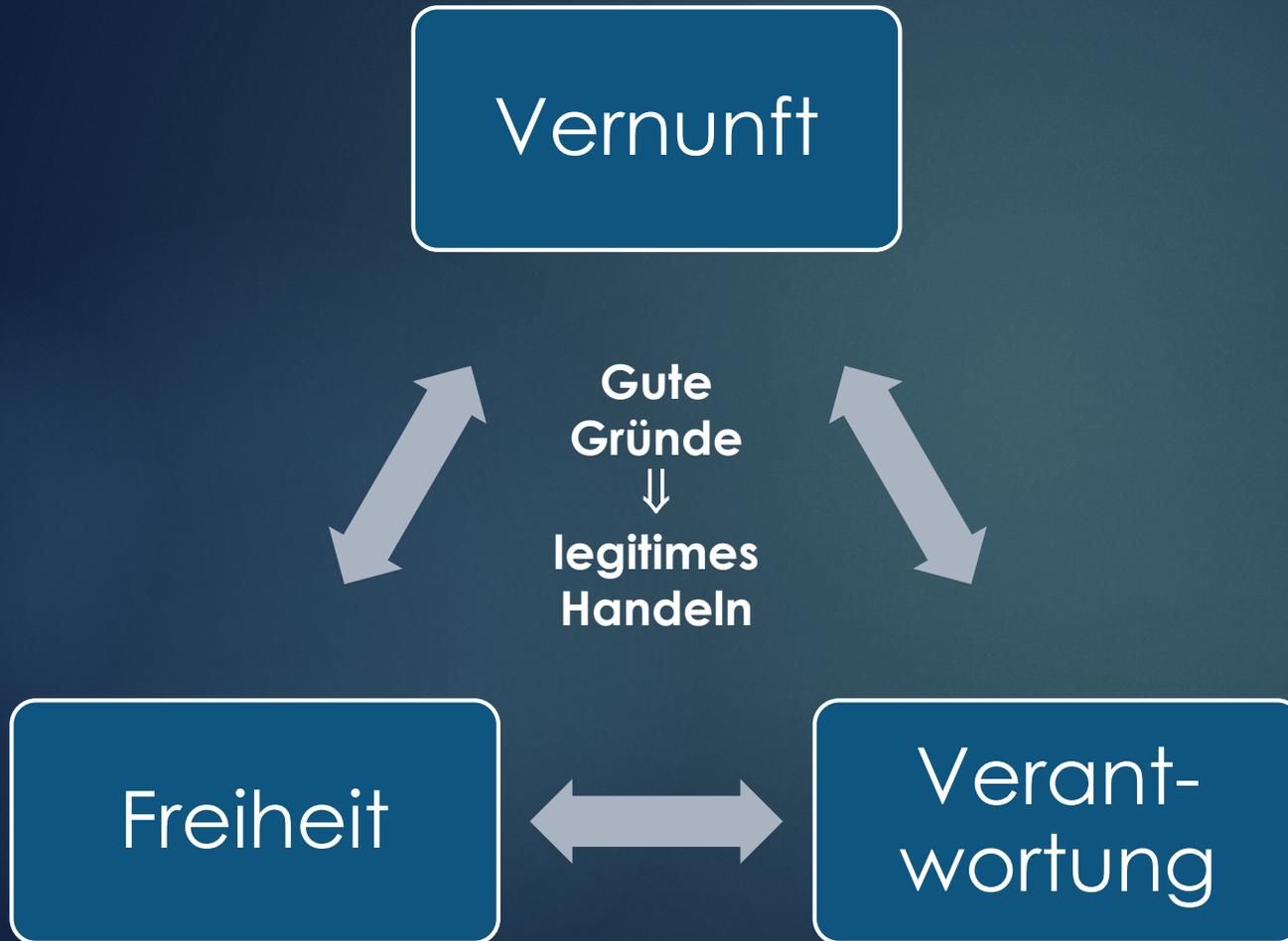
Wollen

Kann ich (jede Person) es wollen, dass von Menschen **geschaffene Intelligenz** den Planeten bevölkert?

Wirken

Kann KI dazu beitragen, **Glück** (Wohlstand) auf dem Planeten zu vermehren?

Kritischer Humanismus



<https://executiveacademy.at/de/news/detail/digitaler-humanismus-fuer-eine-technologie-die-dem-menschen-dient>

Digitaler Humanismus



- ▶ Digitale Technologien sind Werkzeuge, keine Interaktionspartner.
- ▶ Intelligente Maschinen können menschliches Denken und Fühlen simulieren, sie denken und empfinden nicht wie Menschen.
- ▶ KI kann keine Verantwortung tragen.
- ▶ **Ethischer Imperativ des Digitalen Humanismus:**
Welche digitalen Technologien tragen in welchen Anwendungskontexten zu einer nachhaltigen Humanisierung der Lebens- und Arbeitsverhältnisse des Menschen bei?



Literaturhinweise



<https://standards.ieee.org/wp-content/uploads/import/documents/faqs/gieais-faq-11.22.2020.pdf>
<https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>

