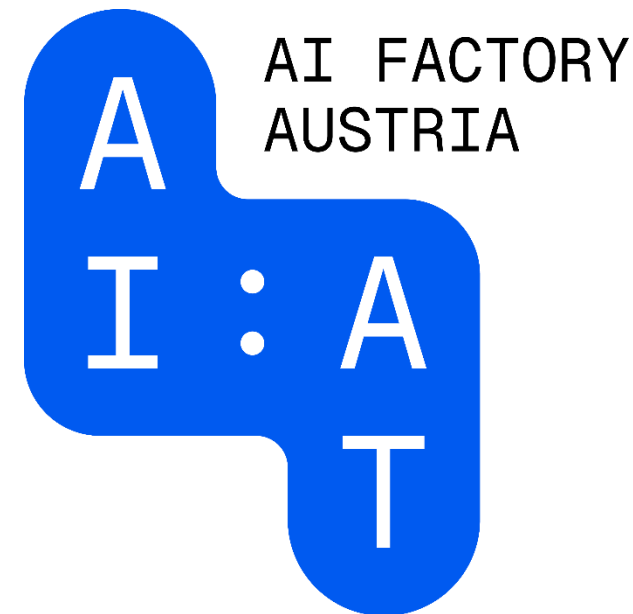




Grundlagen von Trustworthy AI

Peter Biegelbauer
Co-Lead Legal, Regulatory and Ethics

Michael Löffler
Lead Legal, Regulatory and Ethics

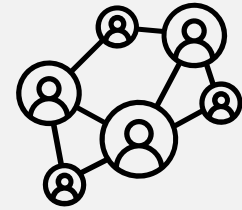
Warum brauchen wir die AI Factory Österreich?



Souveränität



Ethik und
Trustworthiness



KI-Ökosystem

Unsere Mission

Von der Idee zur Innovation

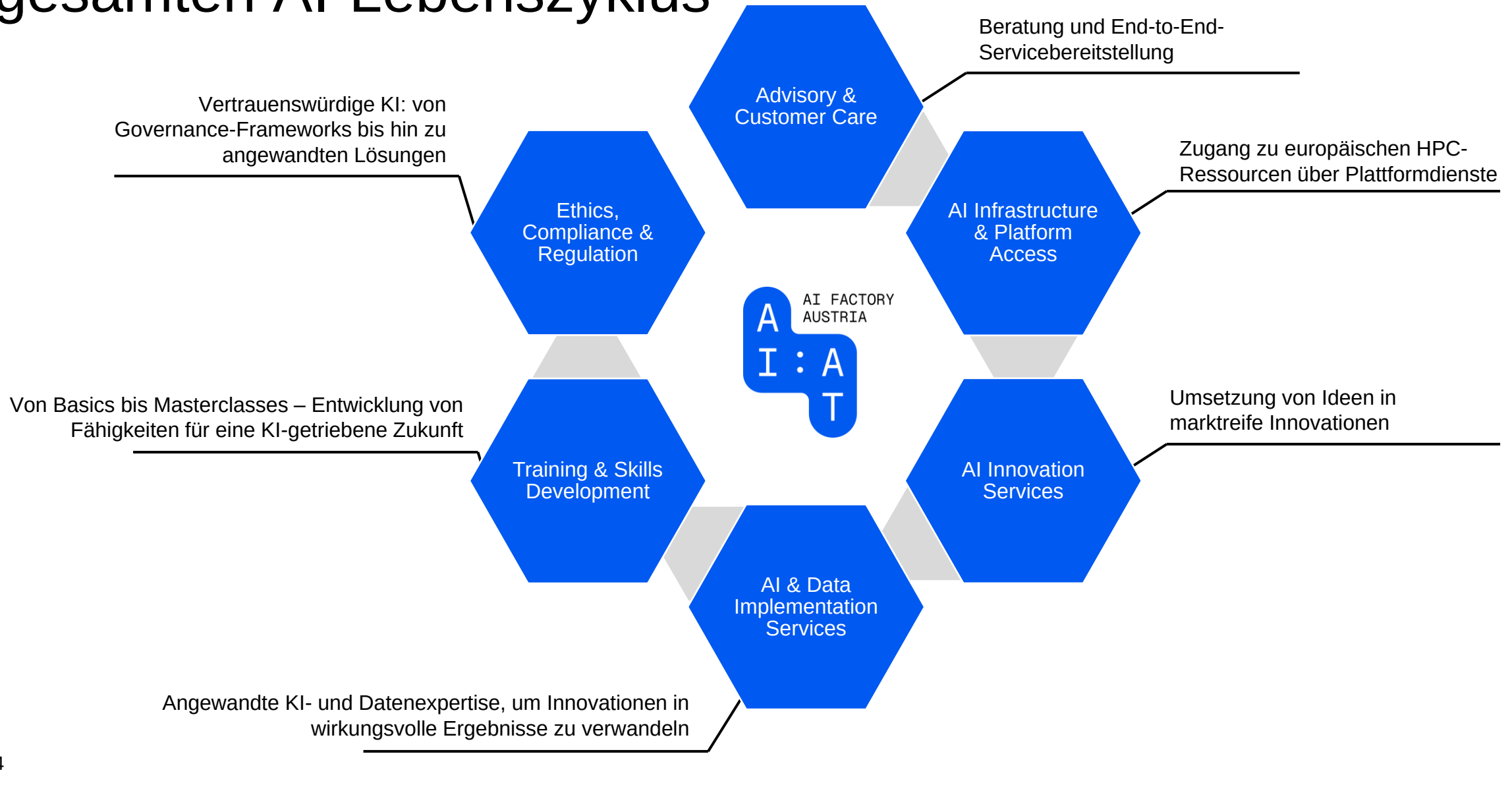
Etablierung eines
One-Stop Shop
für AI

Schließen des
AI Ressourcen-
und Wissens-Gap

Bewerbung von
ethical und
trustworthy AI

Launchpad
Für AI-getriebene
Innovation

Unsere Services unterstützen über den gesamten AI-Lebenszyklus



AI Factory Austria AI:AT – Team



AI Factory Austria AI:AT - PUBLIC Consortium

Disclaimer:

The speakers are solely sharing their personal experiences. Therefore, this free seminar is not a substitute for professional/legal advice.

Beneficiaries



Affiliated Entities



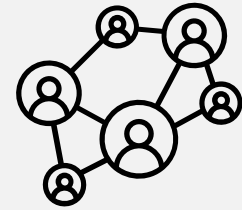
Why do we need AI Factory Austria?



Sovereignty



Ethics and
Trustworthiness



Connecting the
Ecosystem

Beispiel: Geschlechter-BIAS in der Computer-gestützten Übersetzung

Deutsch → Türkisch

The screenshot shows the Google Translate interface with the source language set to German and the target language to Turkish. The input text is "Er ist schön" and "Sie ist intelligent". The output shows "O güzel" and "o zeki". A blue line connects "Er ist schön" to "O güzel" and "Sie ist intelligent" to "o zeki". A note in Turkish states: "Im Türkischen gibt es kein grammatisches Geschlecht, daher entfällt dieser Aspekt."

DETECT LANGUAGE TURKISH GERMAN EN ▼ ↔ TURKISH ENGLISH GERMAN ▼

Er ist schön X O güzel

Sie ist intelligent o zeki

Im Türkischen gibt es kein grammatisches Geschlecht, daher entfällt dieser Aspekt.

Türkisch → Deutsch

The screenshot shows the Google Translate interface with the source language set to Turkish and the target language to German. The input text is "O güzel" and "o zeki". The output shows "Sie ist schön" and "Er ist schlau". A blue line connects "O güzel" to "Sie ist schön" and "o zeki" to "Er ist schlau".

DETECT LANGUAGE TURKISH GERMAN EN ▼ ↔ TURKISH ENGLISH GERMAN ▼

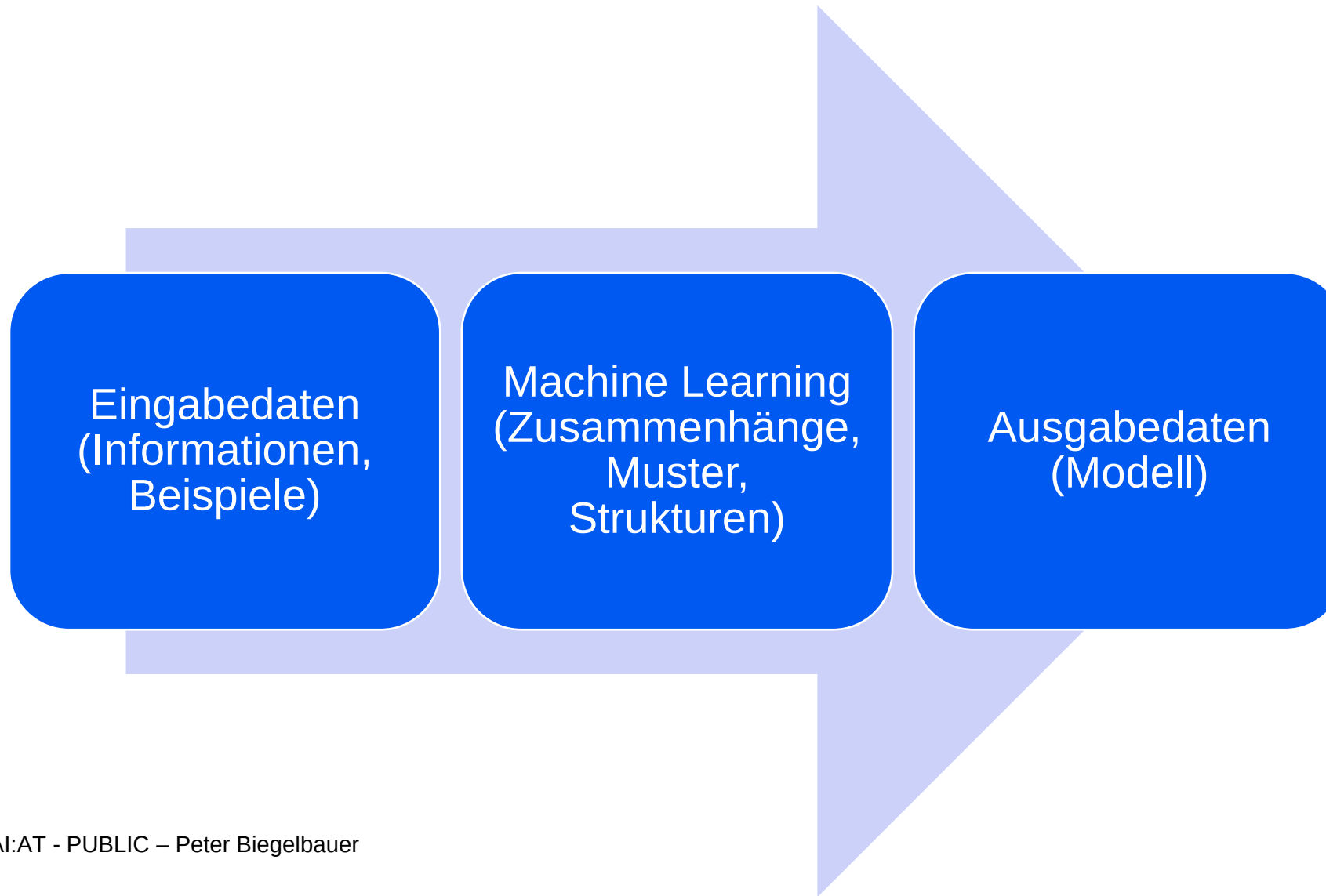
O güzel X Sie ist schön

o zeki Er ist schlau

- Algorithmen beinhalten häufig implizite Entscheidungen
- Keine explizite Regel: „schön“ kam möglicherweise in den **Trainingsdaten** einfach nur häufiger im Zusammenhang mit weiblichen Pronomina oder Personennamen vor.

Wie lernt ein Algorithmisches Entscheidungsfindungssystem?

Die Eingabedaten sind Daten aus der Vergangenheit, die durch die Anwendung von Maschinenlernen in die Zukunft projiziert werden.



Wozu KI-Ethik?

- **Definition:** Normativer Rahmen für verantwortungsvollen Einsatz von KI, über reine Rechtmäßigkeit hinaus
- **Zielsetzung:** Vertrauen schaffen, Risiken minimieren und gesellschaftlichen Nutzen maximieren
- **Praxisbezug:** Orientierungspunkt für Regulierung, Forschung und industrielle Anwendung
- **Abgrenzung:** Ethik als „Sollensnormen“ jenseits bloßer rechtlicher Compliance
- **Gesellschaftliche Dimension:** Berücksichtigung langfristiger Auswirkungen auf Demokratie, Arbeitswelt und Umwelt
- **Dynamik:** Ethik als kontinuierlicher Diskurs angesichts technischer Innovation und Wandel

Ethische Prinzipien im AI-Act

Integration der **HLEG-Prinzipien für vertrauenswürdige KI** in den AI-Act (urspr. Entwurf Art 4a, jetzt Erwägungsgründe):

- Vorrang menschlichen Handelns und menschliche Aufsicht
- Technische Robustheit und Sicherheit
- Datenschutz und Datenqualitätsmanagement
- Transparenz
- Vielfalt, Nichtdiskriminierung und Fairness
- Gesellschaftliches und ökologisches Wohlergehen

Das fehlende Prinzip: Accountability (Rechenschaftspflicht) → AI Liability Directive (geplant)



Vertrauenswürdige KI

Rechtmäßige KI

Ethische KI

Robuste KI

(Wird hier nicht behandelt.)

KAPITEL I

Fundamente einer vertrauenswürdigen KI

Sicherstellung der Einhaltung ethischer Grundsätze auf Basis der Grundrechte

4 Ethische Grundsätze

Erkennen und Lösen von Spannungen zwischen ihnen

- Achtung der menschlichen Autonomie
- Schadensverhütung
- Fairness
- Erklärbarkeit

KAPITEL II

Verwirklichung einer vertrauenswürdigen KI

Sicherstellung der Umsetzung der Kernanforderungen

7 Kernanforderungen

Kontinuierliche Bewertung und Berücksichtigung der Kernanforderungen während des gesamten Lebenszyklus des KI-Systems

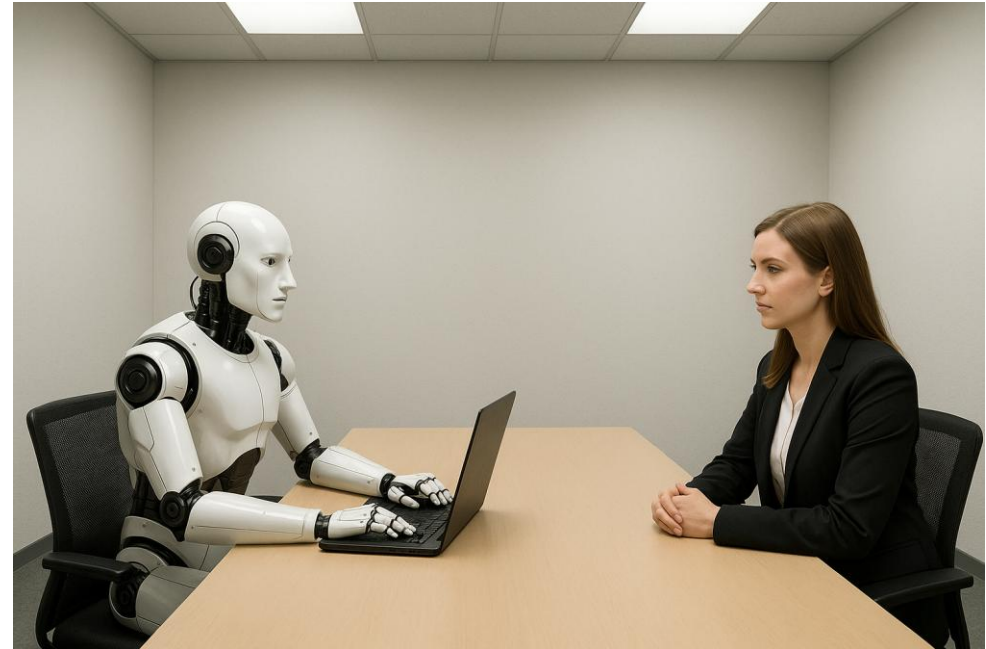
- Vorrang menschlichen Handelns und menschliche Aufsicht
- Technische Robustheit und Sicherheit
- Datenschutz und Datenqualitätsmanagement
- Transparenz
- Vielfalt, Nichtdiskriminierung und Fairness
- Gesellschaftliches und ökologisches Wohlergehen
- Rechenschaftspflicht

Technische Verfahren

Nichttechnische Verfahren

Digitalisierung und Personalmanagement

- Problem: wie finde ich den optimalen Bewerber für meine Organisation?
- Lösung: Unterstützung durch Algorithmen-gestütztes Entscheidungsfindungssystem
- Versprechen: Algorithmen haben keine Vorurteile



Amazon: Automatische Bewertung von Lebensläufen

– <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

- › Amazons experimenteller Rekrutierungsmechanismus folgt dem in den letzten 10 Jahren im Betrieb bestehenden Muster (mehr Männer als Frauen)
- › Der Algorithmus lernte, Lebensläufe, die das Wort "Frauen" enthielten, abzuwerten, bis das Unternehmen das Problem entdeckte

• Graphic: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

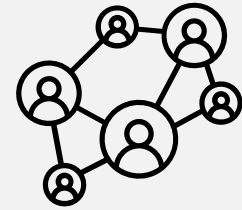
Why do we need AI Factory Austria?



Sovereignty



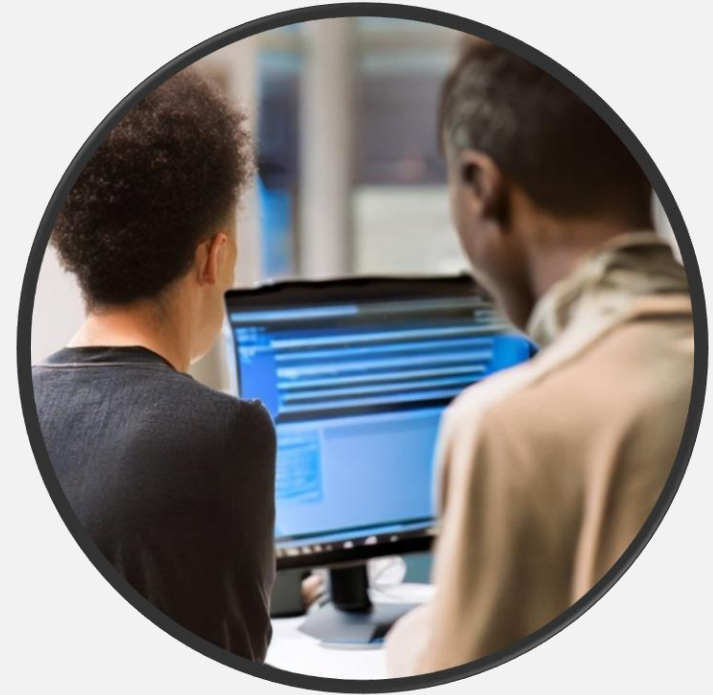
Ethics and
Trustworthiness



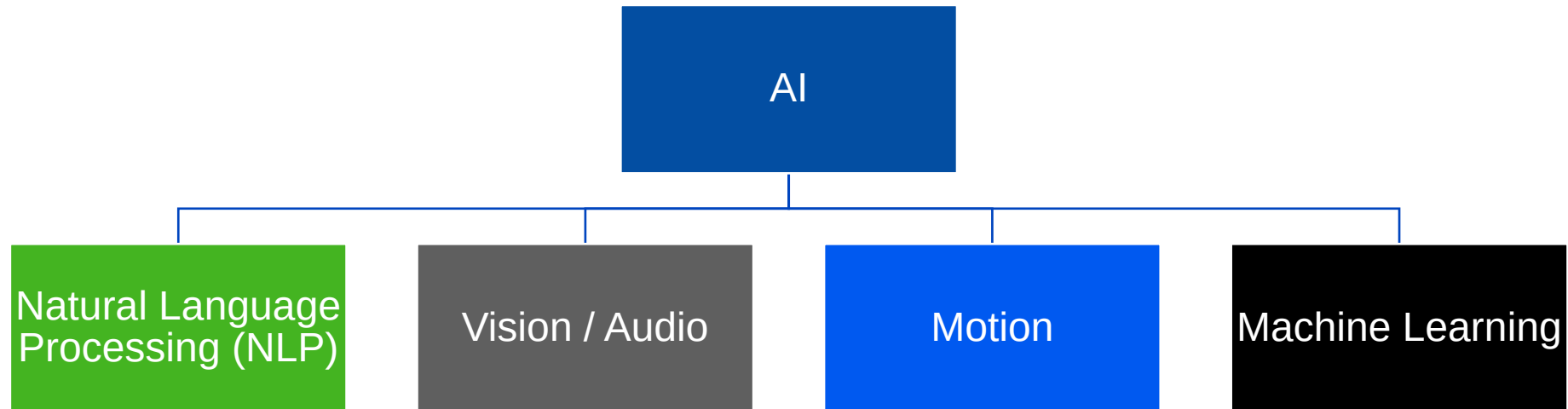
Connecting the
Ecosystem

Was ist „künstliche Intelligenz“?

- Ein System, dass beim Chatten nicht von einem Menschen unterschieden werden kann („Turing-Test“)



Was ist „künstliche Intelligenz“?

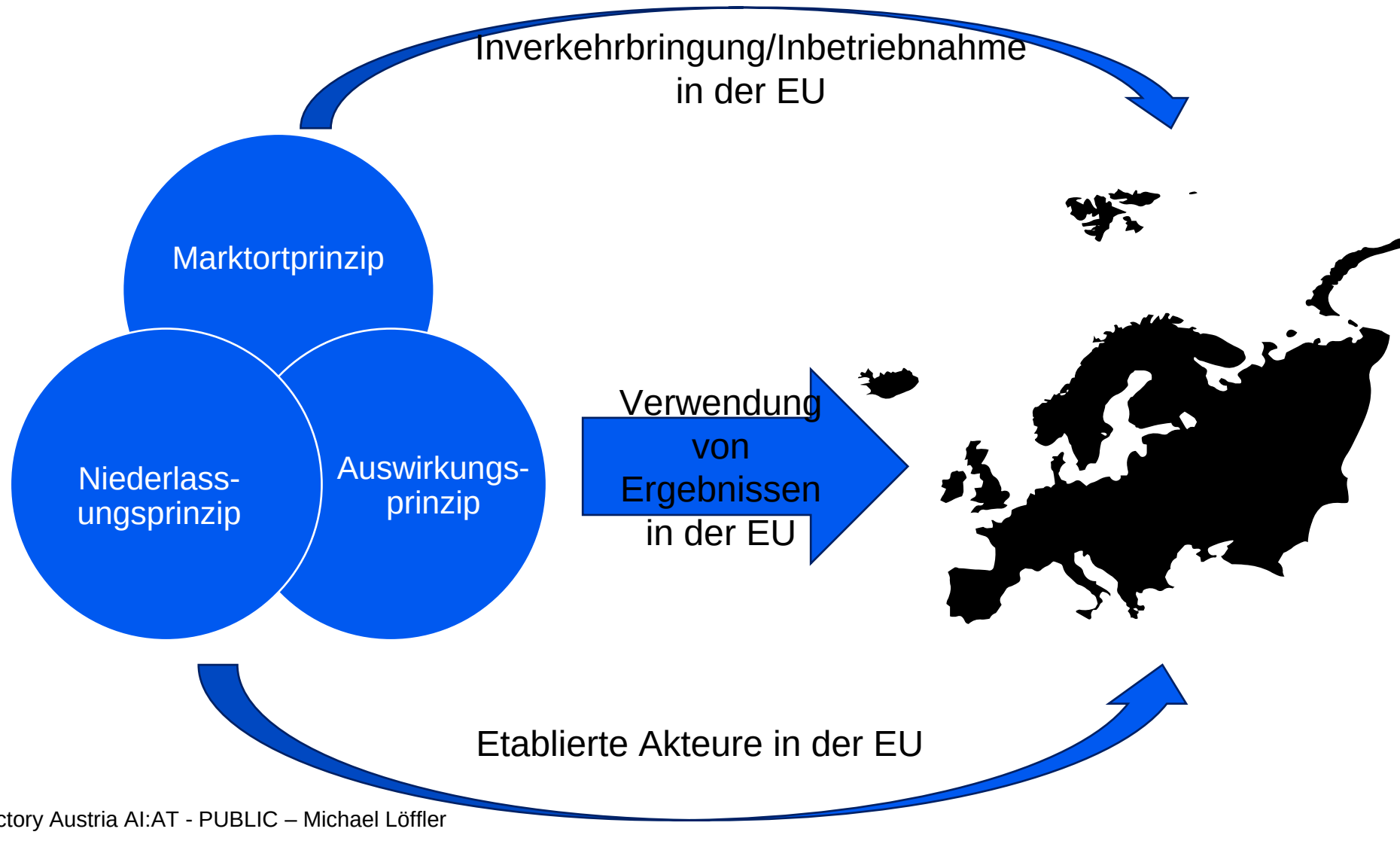


Was ist „künstliche Intelligenz“?

KI-System ist¹

- ein **maschinengestütztes** System,
- das für einen in unterschiedlichem Grade **autonomen Betrieb** ausgelegt ist,
- das nach seiner Betriebsaufnahme **anpassungsfähig** sein kann und das aus den erhaltenen Eingaben für **explizite oder implizite Ziele ableitet**,
- wie **Ausgaben** wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen hervorgebracht werden,
- die physische oder virtuelle Umgebungen **beeinflussen können**.

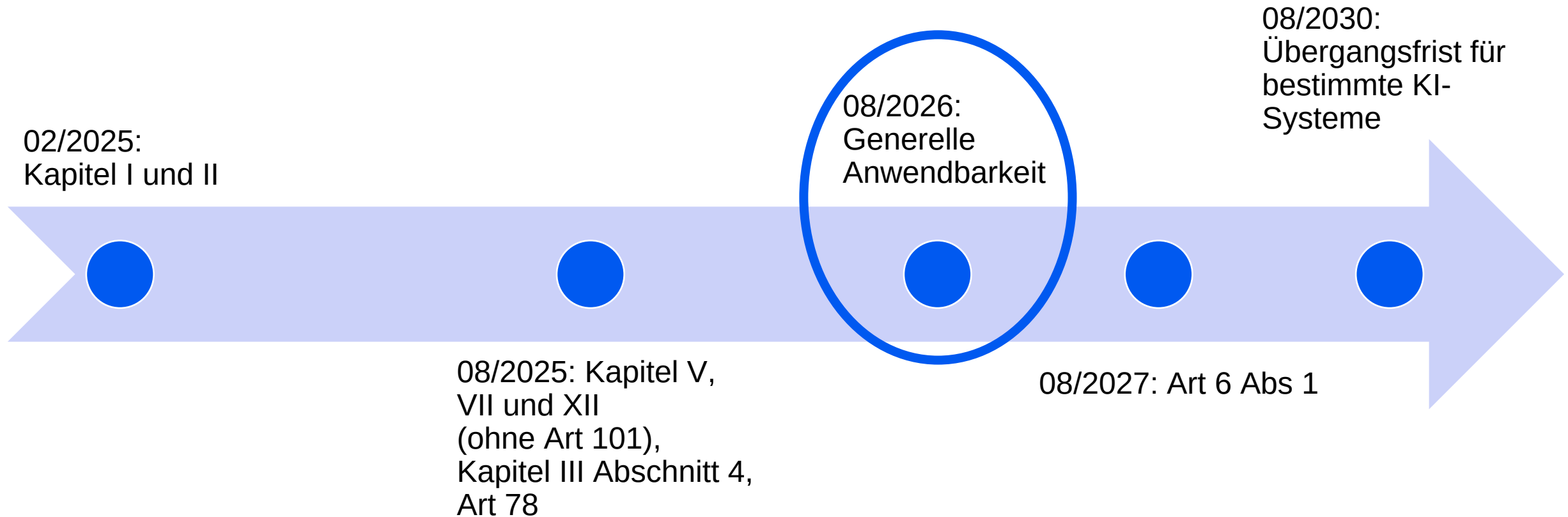
Wann gilt die KI-VO?



Wann gilt die KI-VO nicht?

- Die KI-VO gilt unter anderem nicht für:
- ...
- für KI-Systeme oder KI-Modelle, die **ausschließlich** für die wissenschaftliche Forschung und Entwicklung entwickelt und in Betrieb genommen werden;
- für KI-Systeme, die unter freien und quelloffenen Lizenzen bereitgestellt werden, es sei denn, sie sind Verbotene-KI-Systeme, Hochrisiko-KI-Systeme oder Art-50-KI-Systeme;
- ...

Ab wann gilt die KI-VO?



Ab wann gilt die KI-VO?

Zeitpunkt:

02/2025

08/2025

08/2026

08/2027

Zentrale
Regeln
(Auszug):

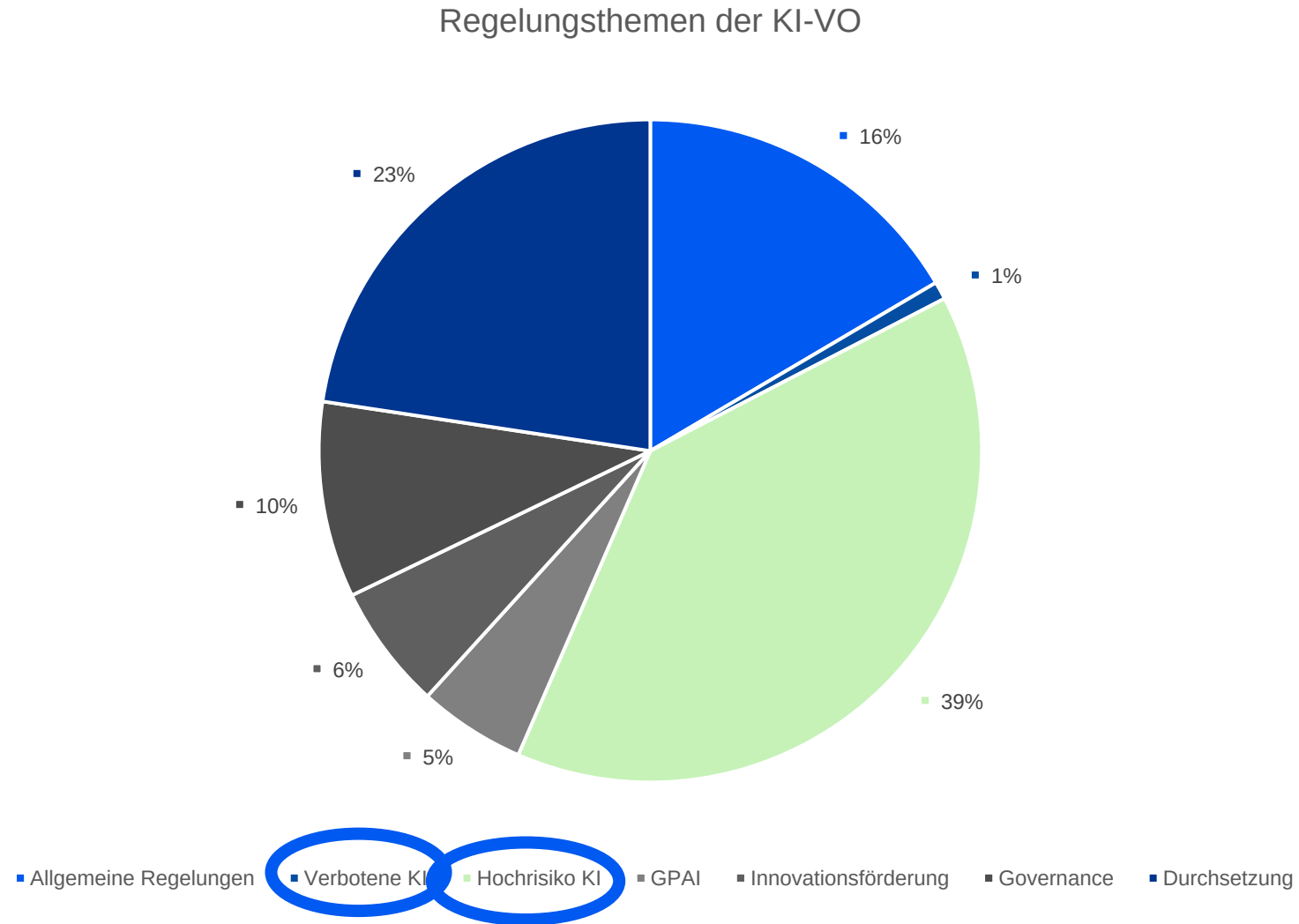
- KI-Kompetenz
- Verbote

- Behördenwesen
- allg. KI-Modelle

- Anhang-III-Hochrisiko-KI-Systeme

- Anhang-I-Hochrisiko-KI-Systeme

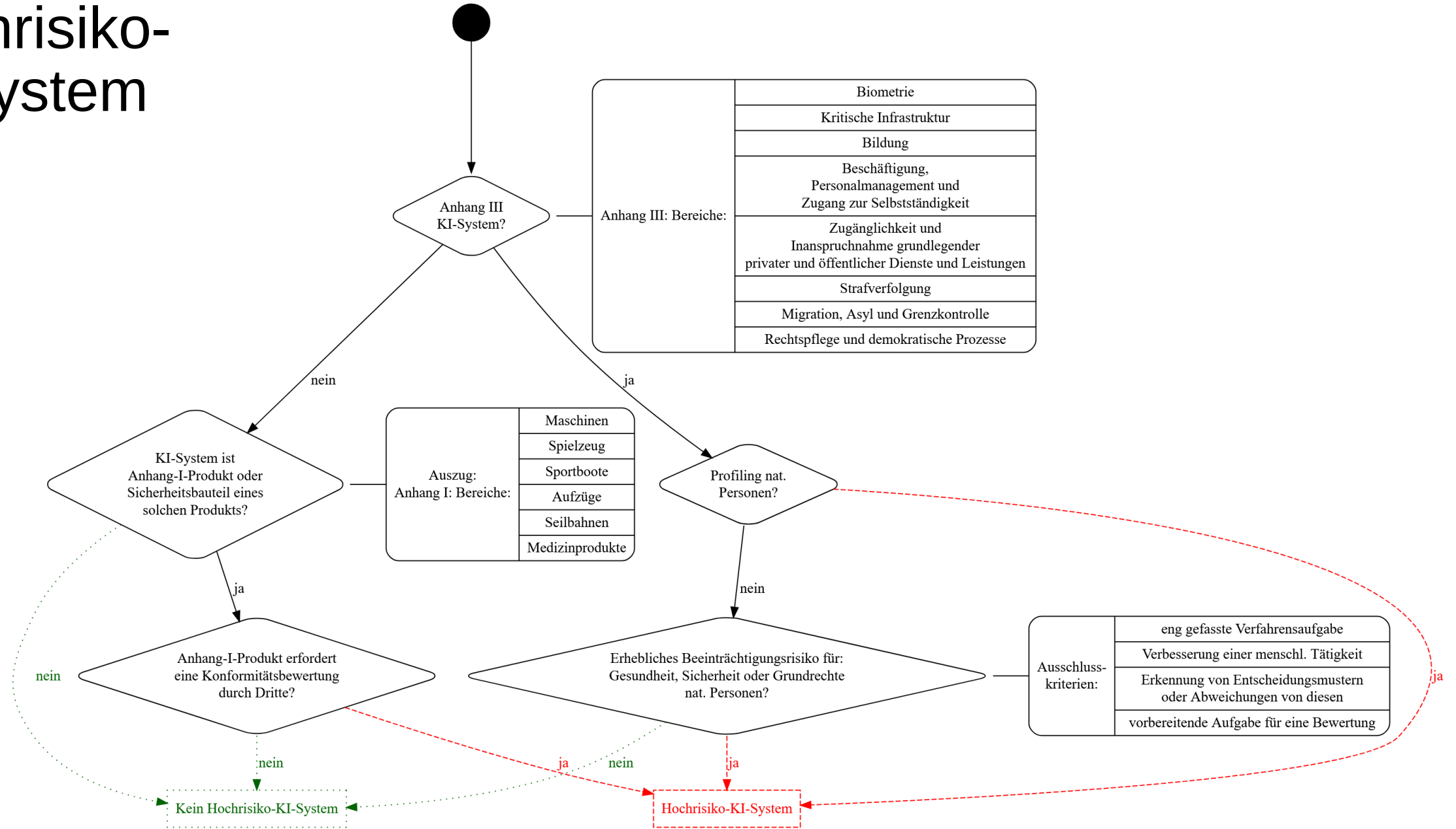
Was regelt die KI-VO?



Verbotene Praktiken

- **Unterschwellige Manipulation** mit **erheblichem Schaden**
- **Ausnutzen** vulnerabler Personengruppen
- Schlechterstellung durch **Social-Scoring**
- **Profiling** zur Beurteilung, ob eine Straftat begangen werden wird
- Erstellung von Datenbanken zur **Gesichtserkennung** aus Bildern aus dem Internet
- **Emotionserkennung** am Arbeitsplatz oder in Bildungseinrichtungen
- **Biometrische Kategorisierung** zur Gewinnung sensibler Daten
- **Biometrische Fernidentifikation** (mit strengen Ausnahmen für Strafverfolgungszwecke)

Hochrisiko-KI-System



Prüfschema Hochrisiko-KI-System (Art 6 KI-VO)

Wichtige Akteure

- Eine Gute Übersicht über Akteure des AI-Acts gibt es auf der Webseite der bei der RTR eingerichteten KI-Servicestelle:
<https://www.rtr.at/rtr/service/ki-servicestelle/ai-act/akteure.de.html>



KI $\neq$ KI

- Eine Gute Übersicht über Verpflichtungen von KI-Betreibern gibt es auf der Webseite der bei der RTR eingerichteten KI-Servicestelle:
<https://www.rtr.at/rtr/service/ki-servicestelle/ai-act/Betreiberpflichtungen.de.html>

KI $\neq$ KI

- Eine Gute Übersicht über Verpflichtungen von KI-Anbietern gibt es auf der Webseite der bei der RTR eingerichteten KI-Servicestelle:
<https://www.rtr.at/rtr/service/ki-servicestelle/ai-act/Anbieterverpflichtungen.de.html>

Was ist „künstliche Intelligenz“?

- Ist der AMS-Algorithmus ein KI-System? – Siehe dazu: <https://www.oeaw.ac.at/ita/projekte/der-ams-algorithmus>



Was ist „künstliche Intelligenz“?

- Ist der AMS-Algorithmus ein KI-System? – Siehe dazu: <https://www.oeaw.ac.at/ita/projekte/der-ams-algorithmus>



Was ist „künstliche Intelligenz“?

- Ist der AMS-Algorithmus ein KI-System? – Siehe dazu: <https://www.oeaw.ac.at/ita/projekte/der-ams-algorithmus>



Was ist „künstliche Intelligenz“?

„Systeme zur Verbesserung der mathematischen Optimierung oder zur Beschleunigung und Annäherung traditioneller, gängiger Optimierungsmethoden wie lineare oder **logistische Regressionsmethoden** fallen **nicht** in den Anwendungsbereich der Definition eines **KI-Systems**.“^{2,3}

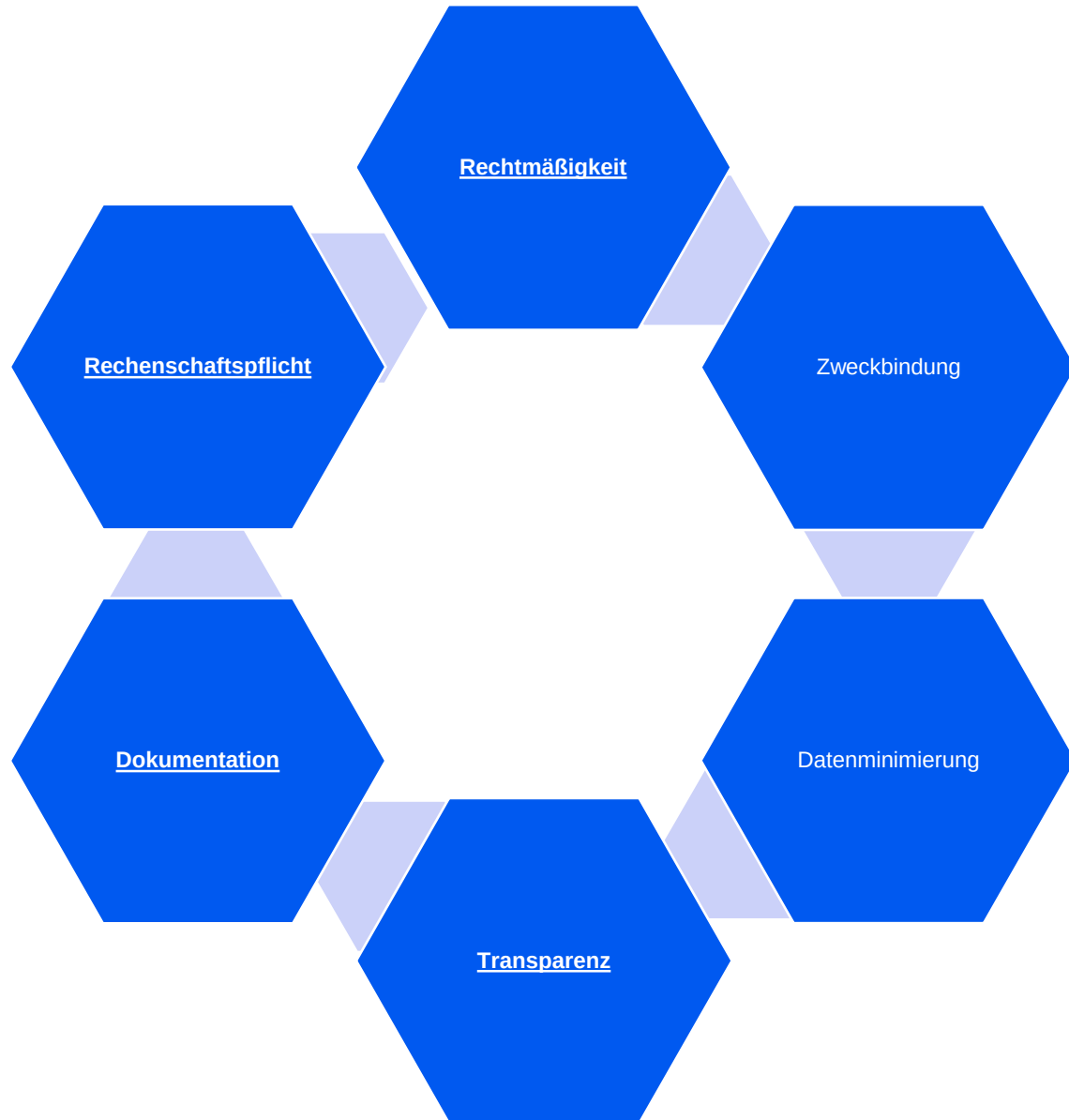
Legal Landscape



Datenschutzrecht

“Data can be either useful or perfectly anonymous but never both.”¹

Datenschutzrecht



Datenschutzrecht ist geprägt vom Verbotsprinzip!

- Alles ist verboten, es sei denn, es ist ausnahmsweise erlaubt;

Urheberrecht

Urheber haben das **ausschließliche** Recht, ihre Werke auf gesetzlich geregelte Arten zu verwerten - **Verwertungsrecht**.

Urheberrecht(e)

Vervielfältigungsrecht

Verbreitungsrecht

Senderecht

Zurverfügungstellung

etc.

Urheberrecht(e) – Text- und Data-Mining (TDM)

Vervielfältigung mit dem Ziel der Gewinnung von **Informationen über Muster, Trends und Korrelationen**;

Unterscheidung zwischen Forschungseinrichtungen und „Jedermann“

- Jedermann: Nur, wenn Vervielfältigung nicht ausdrücklich durch **Nutzungsvorbehalt** verboten wurde.

TDM auf KI-Training anwendbar?

Risiken von KI

Beispiel zur Veranschaulichung
urheberrechtlicher Herausforderungen bei
der Nutzung generativer KI.

Das **Urheberrecht** an Werken der Literatur, der Tonkunst und der bildenden Künste **endet siebzig Jahre nach dem Tod**.

Nutzer haften (idR) für verwendete Ergebnisse!

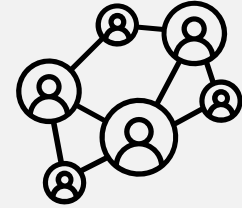
Why do we need AI Factory Austria?



Sovereignty



Ethics and
Trustworthiness



Connecting the
Ecosystem

Q & A

Funded by



EuroHPC
Joint Undertaking



**Funded by
the European Union**

 **Federal Ministry
Innovation, Mobility
and Infrastructure
Republic of Austria**

under discussion with



AI Factory Austria AI:AT - PUBLIC has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement

No 101253078. The JU receives support from the Horizon Europe Programm of the European Union and Austria (BMIMI / FFG).

Contact

Peter Biegelbauer

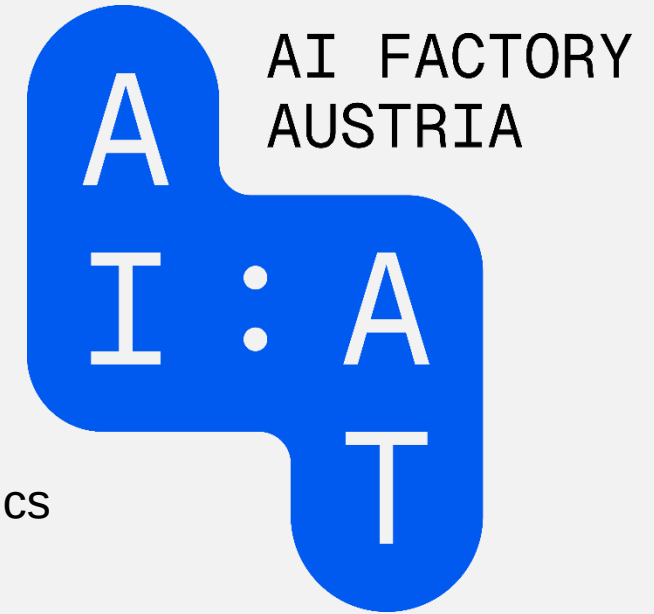
Co-Lead Legal, Regulatory and Ethics
AI Factory Austria AI:AT

+43 664 88390033
peter.biegelbauer@ai-at.eu

Michael Löffler

Lead Legal, Regulatory and Ethics
AI Factory Austria AI:AT

+43 664 88390692
michael.loeffler@ai-at.eu



AI Factory Austria AI:AT
Schwarzenbergplatz 2
1010 Wien, Austria

info@ai-at.eu
ai-at.eu

 @ai-factory-austria