DIHSÜD
Digital Innovation Hub

DIGITALISIERUNG FÜR KMU MÖGLICH MACHEN

DER DIGITAL INNOVATION HUB SÜD ALS KOSTENLOSES SERVICE FÜR KMU

Der DIH SÜD wird unterstützt von:

FFG Forschung wirkt.

Das Land Steiermark

LAND KÄRNTEN

Aktuelle Rechtsthemen
oder
Die neue digitale Ordnung

Die Verordnung zur Künstlichen Intelligenz
KI-VO / AI Act

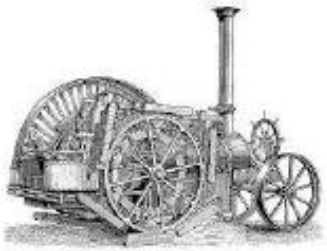
November 2025

Sabine Proßnegg

Basistechnologie

KI als General Purpose Technology (Basistechnologie)

Dampfmaschine



Elektrizität

Computer

Internet / WWW

KI / Generative AI

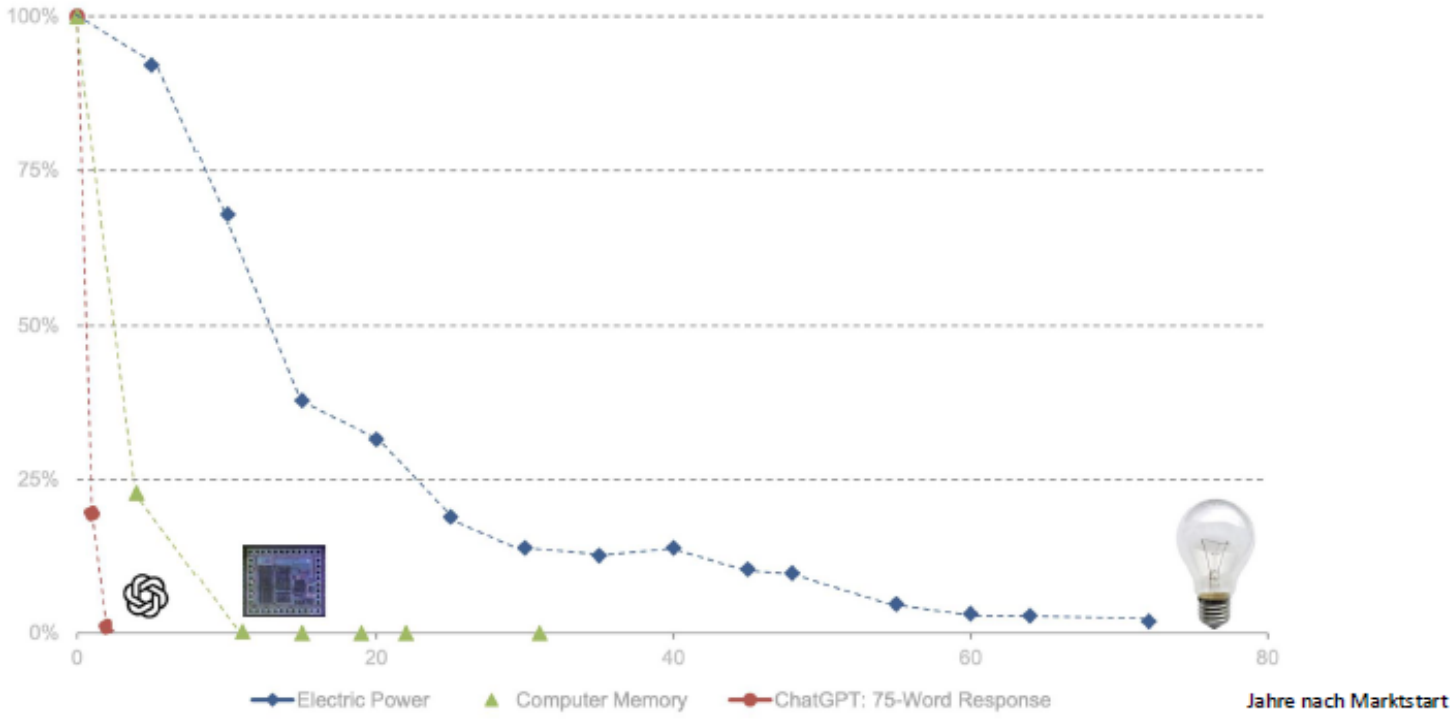
- Breite Anwendbarkeit (✔)
- Kontinuierliche Leistungssteigerung (✔)
- Komplementäre Innovationsfähigkeit (✔)
- Produktivitäts- und Wohlfahrtsimpulse (?)
- Hohe Diffusionsfähigkeit (✔)
- Generelle Zweckoffenheit (✔)

1700 1879 1946 1992 2022

Basistechnologie

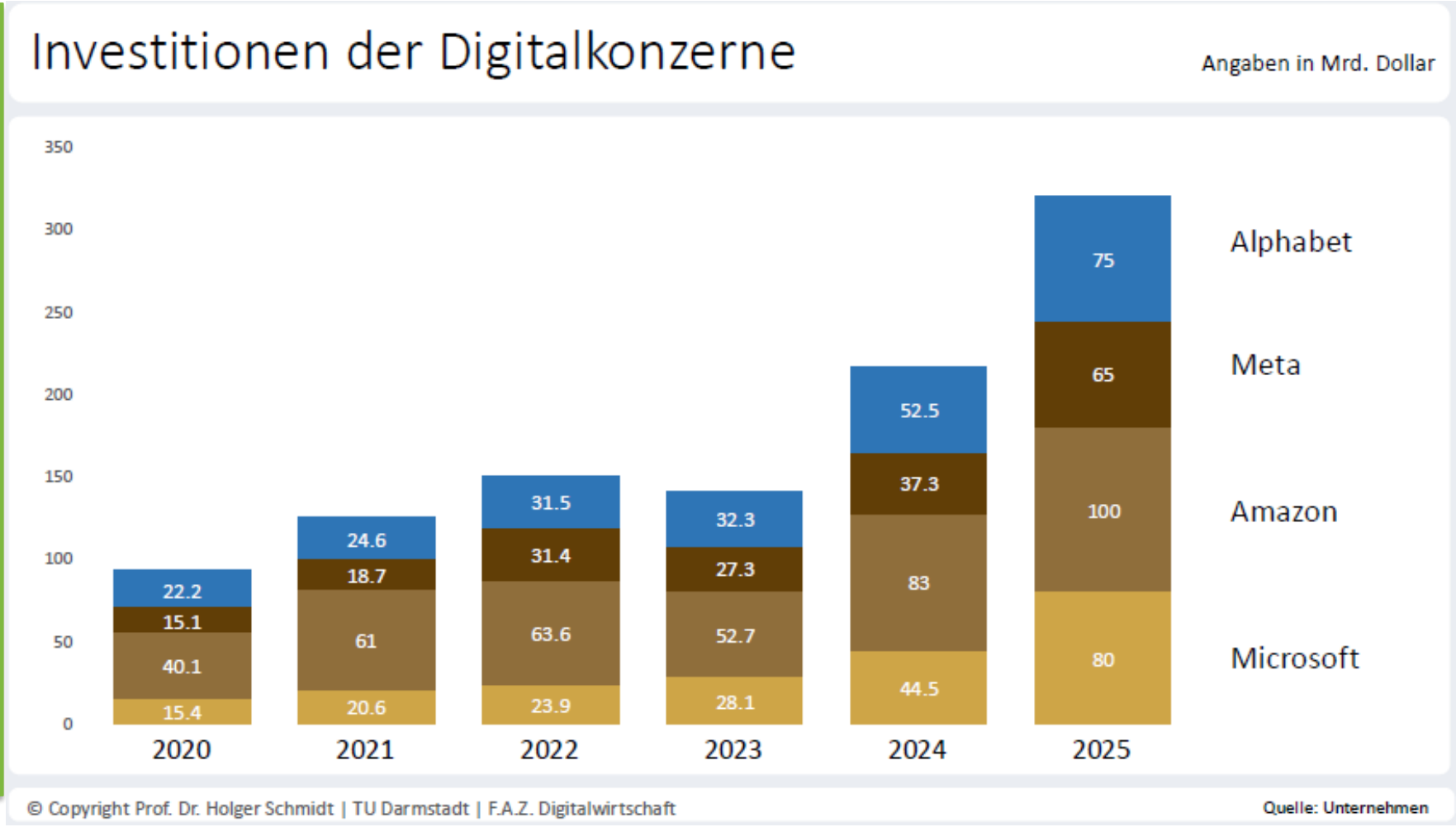
Introduction

Preisverfall der Basistechnologien



Quelle: BOND 2025

Investitionen



Quelle: Unternehmen

KI … und dann?

Der Blick der Ökonomen auf generative KI

Erik Brynjolfsson, Stanford-Ökonom

„Generative KI ist eine der größten und effektivsten Technologien zur Veränderung der Arbeitsweise, die je erfunden wurde.

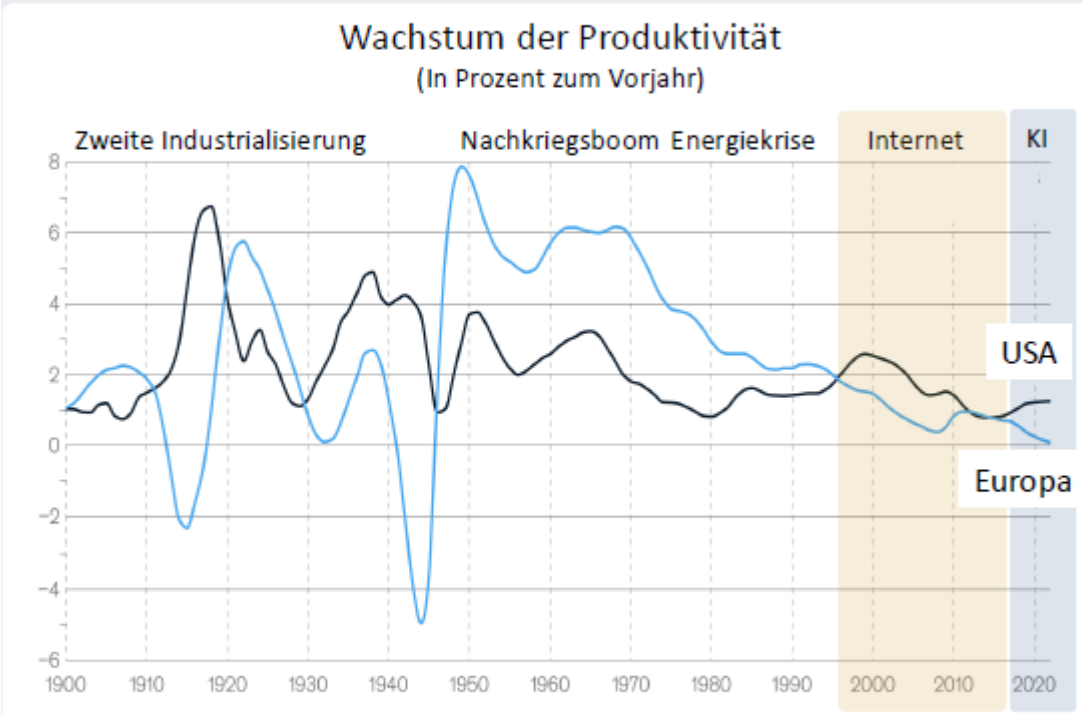
Aber großartige Technologie allein reicht nicht aus. Was man wirklich braucht, ist die Aktualisierung der Geschäftsprozesse, die Umschulung der Belegschaft und manchmal sogar die Änderung der Geschäftsmodelle und der Organisation in großem Umfang."

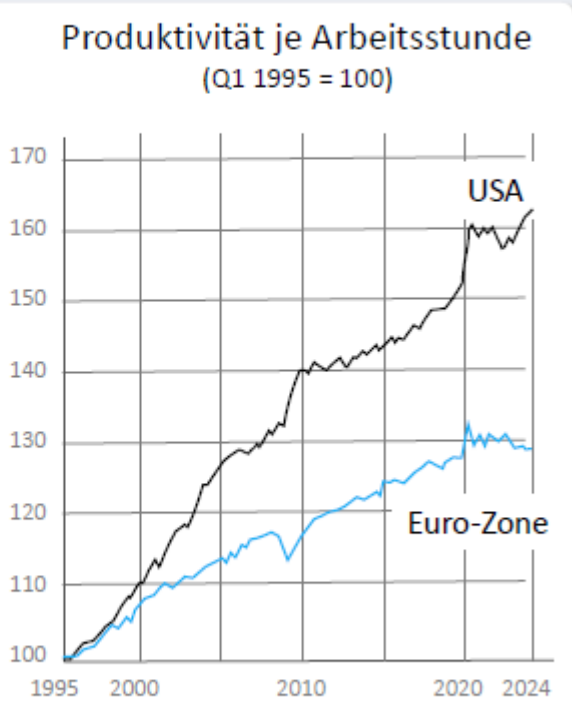
Produktivität

KI-Produktivitätsgewinne in der Finanzbranche

Produktivität

Produktivität in Europa und den USA



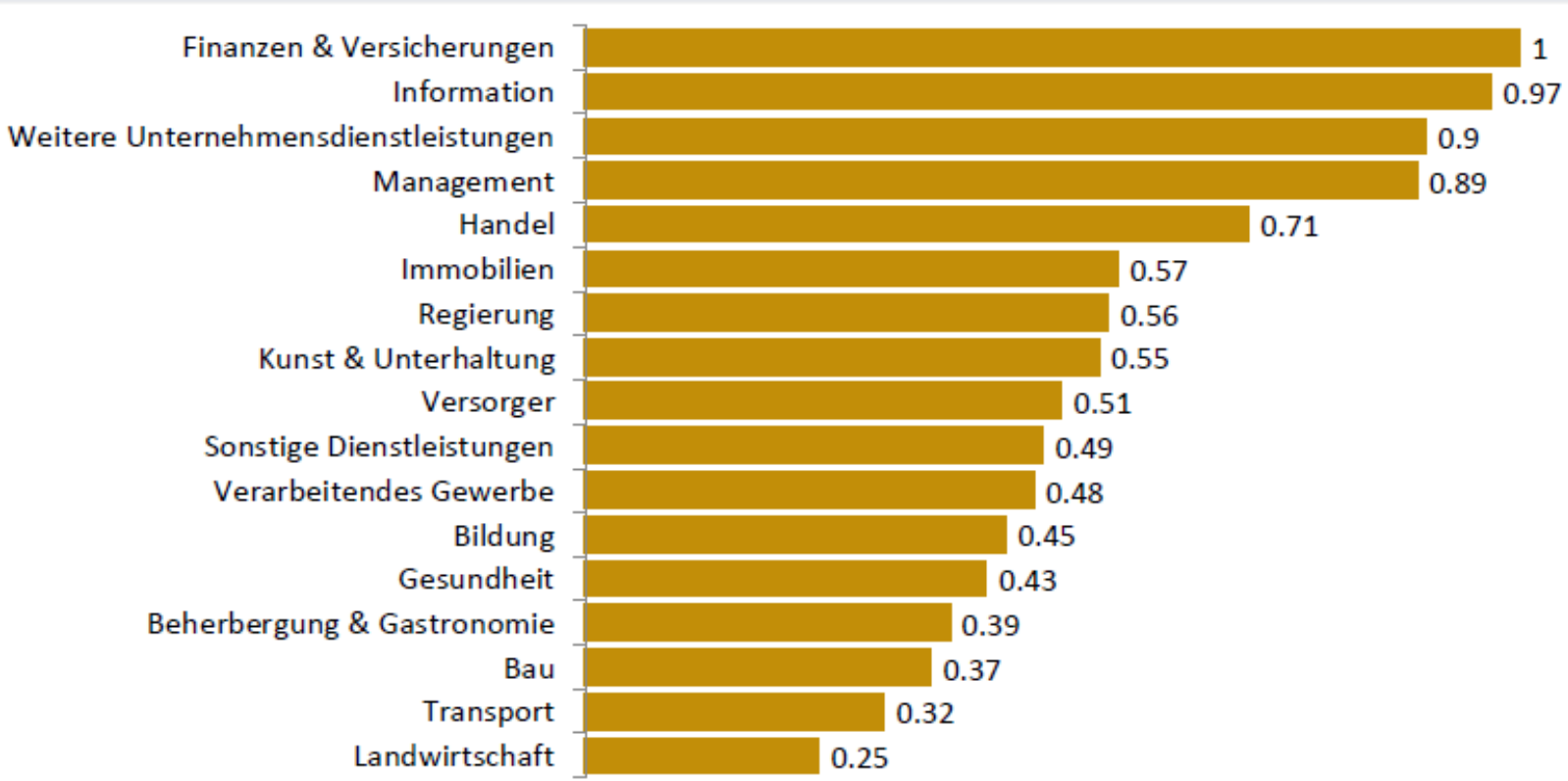


Datenquelle McKinsey 2024 / EZB 2025

KI Effekt

Introduction

Potenzieller GenAI-Effekt nach Branchen



Quelle: Oxford Economics 2024

Neue Kompetenzen

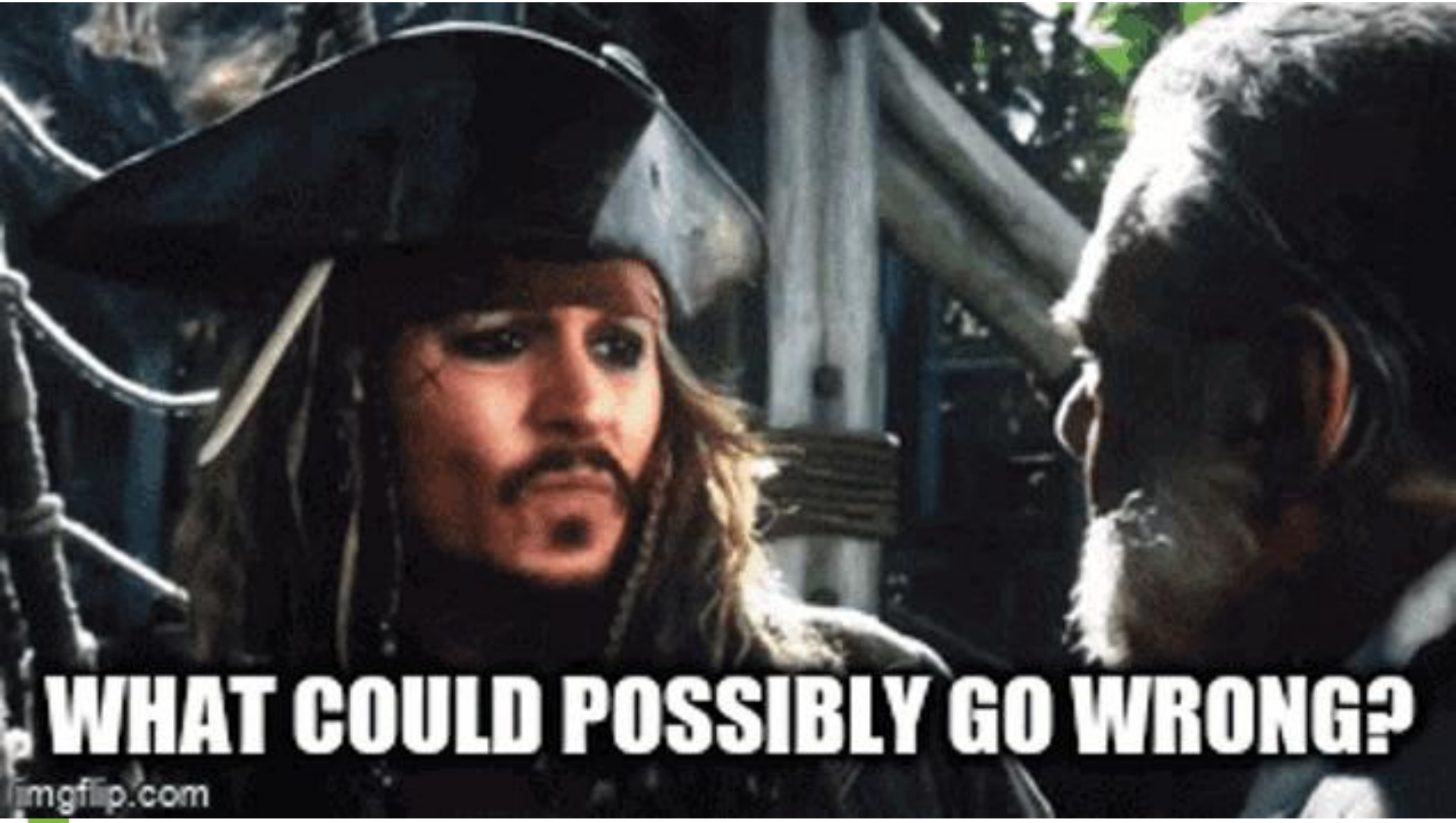
Introduction

Kompetenzen für KI-Einsatz in Unternehmen

Umfrage unter 1003 Unternehmen in Deutschland

KI-Grundwissen		Wissen über die grundlegenden Konzepte und Prinzipien Künstlicher Intelligenz bzw. grundlegendes Verständnis von KI-Systemen.
Praktische generative KI-Kompetenzen	Automatisierung von Prozessen	Gezielter Einsatz von KI, um wiederholbare und regelbasierte Aufgaben zu automatisieren und somit Prozesse effizienter zu gestalten.
	Prompting	Verfassen von spezifischen Anweisungen an ein generatives KI-Modell, um eine gewünschte Ausgabe zu erzeugen.
	Content-Erstellung	Gezielter Einsatz generativer KI für die Produktion von Inhalten wie Texten, Bildern, Videos und Audiodateien.
Praktische Analytische KI-Kompetenzen	Effektive Modell & Tool-Auswahl	Identifikation möglicher KI-Anwendungsfälle und Auswahl der richtigen Modelle und Tools für diese Anwendungsfälle.
	KI-Modellentwicklung	Eigenständige Entwicklung, Schulung und Optimierung von KI-Modellen für spezifische Aufgaben.
	Datengetriebe Entscheidungen	Treffen von Entscheidungen auf Grundlage von Datenanalysen, um Objektivität und Vorhersagegenauigkeit zu verbessern.
	Datenbasiertes Handeln	Verantwortungsbewusstes Arbeiten mit Daten sowie zielgerichteter Einsatz von Daten zur Erkennung und Vermeidung von Risiken.
Kritische und ethische Einordnung der KI		Kritische Bewertung, Analyse und Einordnung der Chancen und Risiken im Zusammenhang mit der Entwicklung und Anwendung von KI-Systemen inklusive einer ethischen Bewertung.

Quelle: Stifterverband 2025

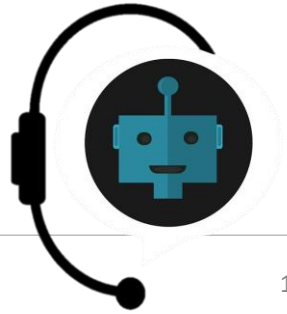

WHAT COULD POSSIBLY GO WRONG?
imgflip.com

FH JOANNEUM
University of Applied Sciences

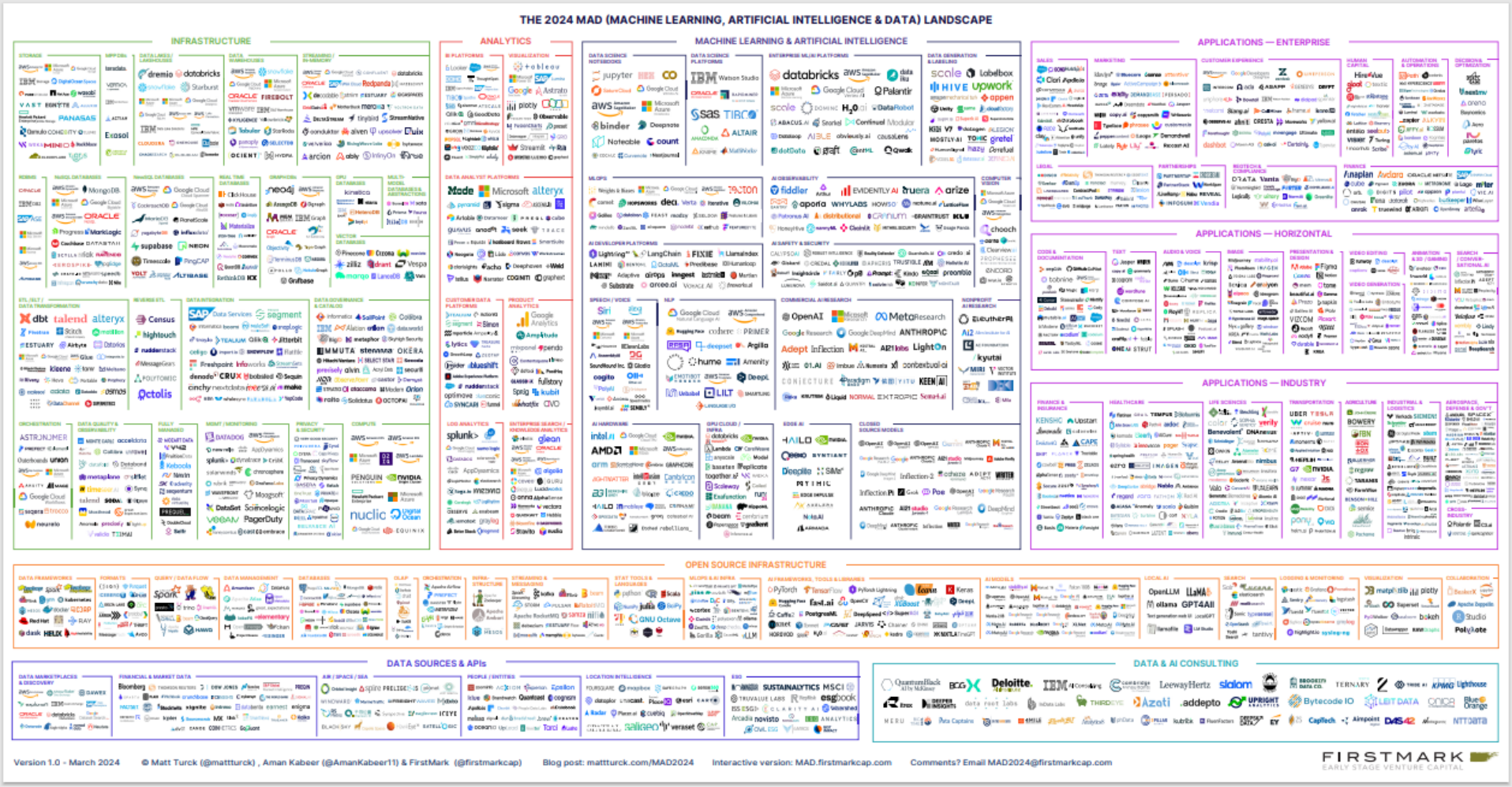
Einleitung und allgemeine Überlegungen

Introduction

- KI verändert derzeit die Gesellschaft -> Chancen und Herausforderungen The AI Act (AIA) (+ GDPR)
- „menschenzentrierte und vertrauenswürdige KI“
- Allerdings gibt es Verbesserungsmöglichkeiten / Unklarheiten
- Der Schwerpunkt muss auf der Praktikabilität liegen etwa bei der menschlichen Aufsicht
- KI-Kompetenz insbesondere bei Nutzern und Betroffenen, samt Abdeckung technologischer, rechtlicher und ethischer Umsetzung



AI landscape



AI in the media

5. Februar 2025
ien 0°C

Die Presse

ACHRICHTEN MEINUNG MAGAZIN Innenpolitik Ausland Klimawandel Economist Finanzen Kultur

Google gibt ausdrückliche Ablehnung von KI-Waffen auf

Nach internen Proteste
aufgenommen, keine KI
Selbstverpflichtung ist
Grundsätze des Untern

tagesschau

Sendung verpasst?

Startseite ▸ Wirtschaft ▸ Digitales ▸ Experten warnen: KI so gefährlich wie Pandemien oder Atomkrieg

BBC

Home News Sport Business Innovation Culture Arts Travel Earth Audio Video Live

ChatGPT can answer questions using natural, human-like language and mimic other writing styles

A New York lawyer is facing a court hearing of his own after his firm used AI tool ChatGPT for legal research.

A judge said the court was faced with an "unprecedented circumstance" after a filing was found to reference example legal cases that did not exist.

The lawyer who used the tool told the court he was "unaware that its content could be false".

ChatGPT creates original text on request, but comes with warnings it can "produce inaccurate information".

Experten warnen vor Risiken

KI so gefährlich wie Pandemien oder Atomkrieg

Stand: 30.05.2023 18:16 Uhr

Führende Experten warnen davor, die Risiken Künstlicher Intelligenz zu unterschätzen. Sie bewerten die Gefahren durch KI ähnlich hoch wie bei Pandemien oder einem Atomkrieg.

SPIEGEL Netzwelt

Volkswagen-Datenpanne: Rückschlüsse auf das Leben der Menschen am Lenkrad Foto: [M] DER SPIEGEL; Fotos: Paul Langrock, Thomas Starck / AUTO BILD, picture alliance / dpa / Audi AG

Datenleck beim Volkswagen-Konzern

Wir wissen, wo dein Auto steht

VW hat mit einer neuen Blamage zu kämpfen. Bewegungsdaten von 800.000 E-Autos sowie Kontaktinformationen zu den Besitzern standen ungeschützt im Netz. Sichtbar war, wer wann zu Hause parkt, beim BND oder vor dem Bordell. Die SPIEGEL-Recherche.

Von **Patrick Beuth**, **Flüpke**, **Max Hoppenstedt**, **Michael Kreil**, **Marcel Rosenbach** und **Rina Wilkin**
27.12.2024, 10.25 Uhr • aus **DER SPIEGEL 1/2025**

DERSTANDARD

Recht International Inland Wirtschaft Web Sport Panorama Kultur Etat Wissenschaft Lifestyle Diskurs

127 Postings

VERWALTUNGSGERICHTSHOF

Entscheidet die KI über Jobs? Höchst lässt AMS-Algorithmus erneut prüfe

Das umstrittene Programm soll Arbeitssuchende nach ihrer Vermittelt einstufen. Jetzt muss geprüft werden, ob am Ende doch ein Mensch e

Jakob Pflügl
10. Februar 2024, 09:00

127 Postings Später lesen

DERSTANDARD

NEWSLETTER KÜNSTLICHE INTELLIGENZ

Deepseek ist effizienter – aber die grüne KI ist noch nicht da

Das chinesische KI-Modell benötigt zwar weniger Rechenleistung als ChatGPT und Co. Doch die Klimabilanz der Technologie hängt trotzdem von vielen anderen Faktoren ab

Kolumne / Philip Pramer
3. Februar 2025, 20:37

3 Postings Später lesen

Erst rund eine Woche ist es her, als kaum jemand die KI-Führerschaft der Vereinigten Staaten infrage stellte. OpenAI, Anthropic, Meta, Nvidia – sie alle schienen der außeramerikanischen Konkurrenz meilenweit voraus.

AI fails

FH | JOANNEUM
University of Applied Sciences

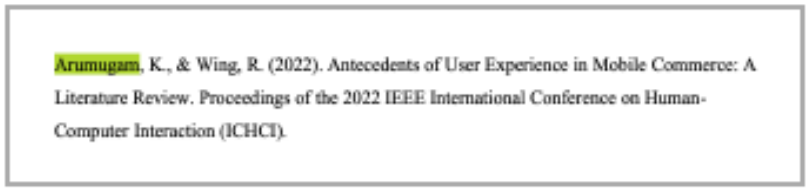
When AI goes wrong: 13 examples of AI mistakes and failures

Air Canada, the largest airline in Canada, was ordered to compensate a passenger who received incorrect refund information from its chatbot. The company acknowledged that the chatbot's response contradicted the airline's policies.

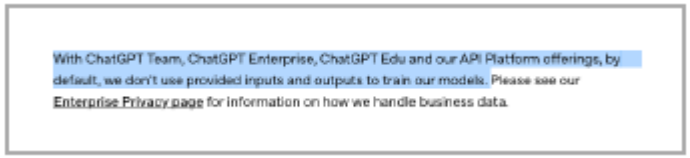
In a New York federal court filing, one of the lawyers was caught citing non-existent legal cases. The attorney had used ChatGPT to conduct legal research, and the AI tool provided fake case references, which the lawyer included in his filing.

The Swedish fintech company Klarna introduced an AI-powered customer support assistant that quickly made a significant impact. Within its first month, the AI handled 2.3 million conversations, equivalent to two-thirds of customer inquiries across 23 markets, supporting 35 languages. One user prompted the chatbot to generate Python code, a task well outside the intended scope of a customer support tool.

KI an der FHJ


Arumugam, K., & Wing, R. (2022). Antecedents of User Experience in Mobile Commerce: A Literature Review. Proceedings of the 2022 IEEE International Conference on Human-Computer Interaction (ICHCI).

Existiert in der Kombination nicht.


With ChatGPT Team, ChatGPT Enterprise, ChatGPT Edu and our API Platform offerings, by default, we don't use provided inputs and outputs to train our models. Please see our Enterprise Privacy page for information on how we handle business data.

Management war ahnungslos
=> KI wurde im Unternehmen verboten.


The exposé is structured to provide a clear overview of:

- The security rationale behind air-gapped systems
- The operational challenges that drive the need for controlled connectivity
- Current technological approaches and solutions
- Key implementation considerations
- Future developments in the field

This should provide a solid foundation for your seminar work while demonstrating understanding of both the technical and strategic aspects of air-gapped wireless connectivity.

Retry

Claude can make mistakes. Please double-check responses.

Der/die Studierende war ahnungslos.

FH JOANNEUM
University of Applied Sciences

Human vs. AI decision making

Pexels.com: Andrea Piacquadio.

Pexels.com: ThisISEngineering.

AI is not neutral, it can make mistakes, needs human interpretation

Example: PlagScan ..% likely to be a plagiarism / AI generated – was heist das konkret für uns?

FH JOANNEUM
University of Applied Sciences

KI-VO – Die Struktur

Structure AIA

- ➔ Die KI-VO hat 13 Kapitel, 113 Artikel, 180 Erwägungsgründe
- ➔ Art 3 der KI-VO enthält 68 Nummern mit Definitionen
- ➔ KI-VO sieht ein Netz an Compliance-, Prüf- und Genehmigungsvorschriften für KI - auch für KMU

  iVm hohe Strafen
- ➔ Art 5 in Verbindung mit Annex II listet verbotene Praktiken auf
- ➔ Die KI-VO ist teilweise vage zB deep fake
- ➔ Art 6 in Verbindung mit Annex I and Annex III Hochrisiko KI

  Prof. Borges: das geplante KI Register ist ein echtes Service/Serviceleistung der EU, eine Liste zur Verfügung zu stellen;

Der AI Act („KI-Verordnung“) wird bis August 2026 in drei Phasen anwendbar. Die erste Phase begann am 2.2.2025 mit der Anwendbarkeit der Kapitel I und II.

Kapitel I enthält in Artikel 4 eine Verpflichtung zur KI-Kompetenz. Artikel 2 regelt, welche KI-Anwendungen ab 2.2.2025 verboten sind.

FH JOANNEUM
University of Applied Sciences

KI-VO – Struktur XIII KAPITEL

Structure AIA

- ➔ KAPITEL I ALLGEMEINE BESTIMMUNGEN / GENERAL PROVISIONS
- ➔ KAPITEL II VERBOTENE PRAKTIKEN IM KI-BEREICH / PROHIBITED AI PRACTICES
- ➔ KAPITEL III HOCHRISIKO-KI-SYSTEME / HIGH RISK SYSTEM
- ➔ KAPITEL IV TRANSPARENZPFLICHTEN FÜR ANBIETER UND BETREIBER BESTIMMTER KI-SYSTEME / TRANSPARENCY OBLIGATIONS FOR PROVIDERS AND DEPLOYERS OF CERTAIN AI SYSTEMs
- ➔ KAPITEL V KI-MODELLE MIT ALLGEMEINEM VERWENDUNGSZWECK / GENERAL-PURPOSE AI MODELS
- ➔ KAPITEL VI MASSNAHMEN ZUR INNOVATIONSFÖRDERUNG / MEASURES IN SUPPORT OF INNOVATION
- ➔ KAPITEL VII GOVERNANCE
- ➔ KAPITEL VIII EU-DATENBANK FÜR HOCHRISIKO-KI-SYSTEME / EU DATABASE FOR HIGH-RISK AI SYSTEMS
- ➔ KAPITEL IX BEOBACHTUNG NACH DEM INVERKEHRBRINGEN, INFORMATIONSAUSTAUSCH UND MARKTÜBERWACHUNG / POST-MARKET MONITORING, INFORMATION SHARING AND MARKET SURVEILLANCE
- ➔ KAPITEL X VERHALTENSKODIZES UND LEITLINIEN / CODES OF CONDUCT AND GUIDELINES
- ➔ KAPITEL XI BEFUGNISÜBERTRAGUNG UND AUSSCHUSSVERFAHREN / DELEGATION OF POWER AND COMMITTEE PROCEDURE
- ➔ KAPITEL XII SANKTIONEN / PENALTIES
- ➔ KAPITEL XIII SCHLUSSBESTIMMUNGEN / FINAL PROVISIONS

FH JOANNEUM
University of Applied Sciences

X Annexes

- ANNEX I Liste der Harmonisierungsrechtsvorschriften der Union/ Harmonisation legislation
- SECTION II Liste der Straftaten gemäß Artikel 5 Absatz 1 Unterabsatz 1 Buchstabe h Ziffer iii / Criminal offences
- ANNEX III Hochrisiko-KI-Systeme gemäß Artikel 6 Absatz 2 / High Risk
- ANNEX IV Technische Dokumentation gemäß Artikel 11 Absatz 1 / Technical documentation TÜV Leitfaden
- ANNEX V EU-Konformitätserklärung / EU declaration of conformity
- ANNEX VI Konformitätsbewertungsverfahren auf der Grundlage einer internen Kontrolle / Conformity assessment procedure based on internal control
- ANNEX VII Konformität auf der Grundlage einer Bewertung des Qualitätsmanagementsystems und einer Bewertung der technischen Dokumentation / Conformity based on an assessment of the quality management system and an assessment of the technical documentation
- ANNEX VIII Bei der Registrierung des Hochrisiko-KI-Systems gemäß Artikel 49 bereitzustellende Informationen / Information to be submitted upon the registration of high-risk AI systems in accordance with Article 49
- ANNEX IX Bezüglich Tests unter Realbedingungen gemäß Artikel 60 bei der Registrierung von in Anhang III aufgeführten Hochrisiko-KI-Systemen bereitzustellende Informationen / Information to be submitted upon the registration of high-risk AI systems listed in Annex III in relation to testing in real world conditions in accordance with Article 60
- ANNEX X Rechtsvorschriften der Union über IT-Großsysteme im Raum der Freiheit, der Sicherheit und des Rechts / Union legislative acts on large-scale IT systems in the area of Freedom, Security and Justice

FH JOANNEUM
University of Applied Sciences

KI-VO

Ziel der KI-VO

Kein umfassender Rechtsrahmen, sondern Produktsicherungsrecht für KI, ergänzend für bisherige Verbote, Transparenz und Individualrechtsschutz Vorgaben.

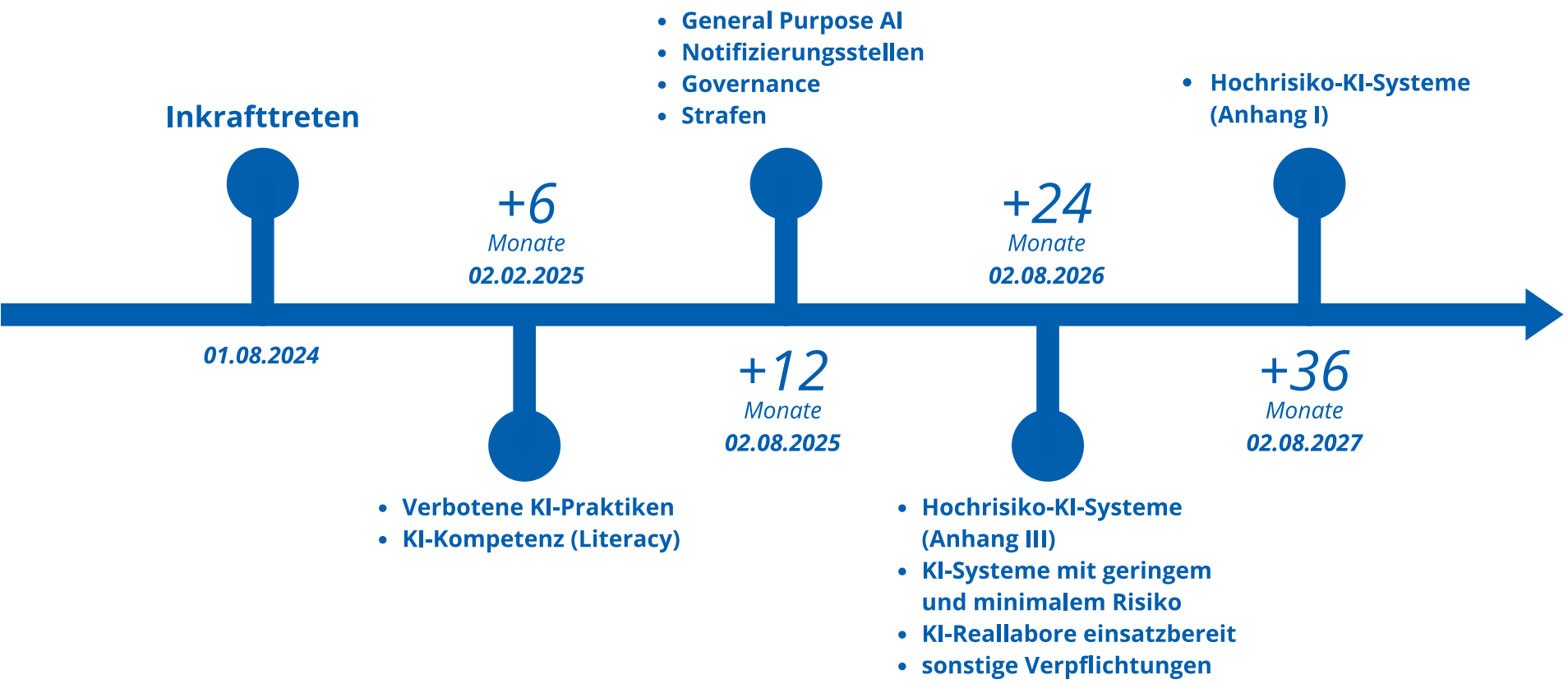
Starker Fokus auf staatliche Durchsetzung
Art 64ff AI-Act: AI Office, AI Board, Advisory Forum, Scientific Panel, Notifying Authority, Market Surveillance Authority;

Passierschein A38? - Die Zuständigkeitsregelungen des AI Acts im nationalen und supranationalen Behördendschungel

Thorsten Ammann; Florian Achnitz; Ludwig Lauer, 21.11.2024

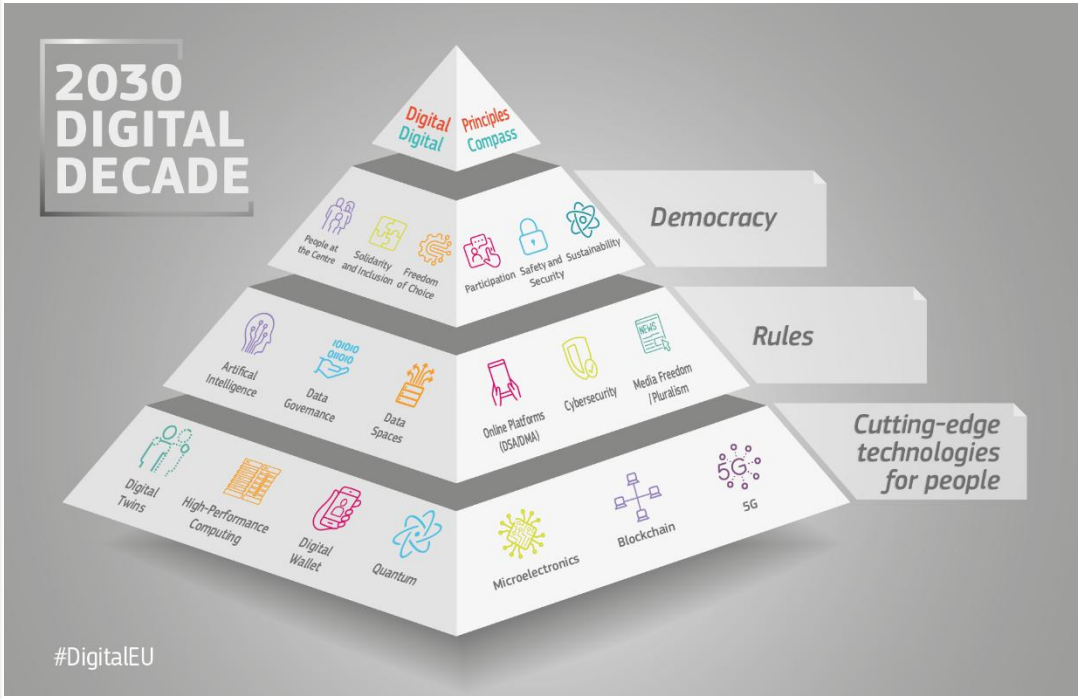
AI Act: Zeitlicher Rahmen

Überblick über die wichtigsten Bestimmungen, die erst nach und nach Gültigkeit erlangen



FH JOANNEUM
University of Applied Sciences

EU Digitalisierungsstrategie 2030



Die EU strebt eine „menschenzentrierte und vertrauenswürdige KI" an.

Wie sieht das in der Praxis aus?

Die KI-Governance muss wirtschaftliche Interessen und das Wohlergehen der Menschen in Einklang bringen.

FH JOANNEUM
University of Applied Sciences

Ö KI Strategie AI Mission Austria

KI-Strategie - Artificial Intelligence Mission Austria

FH JOANNEUM
University of Applied Sciences

KI-VO Art 1 Abs 1 KI-VO

Markt

Artikel 1

Gegenstand

(1) Zweck dieser Verordnung ist es, das Funktionieren des Binnenmarkts zu verbessern und die Einführung einer auf den Menschen ausgerichteten und vertrauenswürdigen künstlichen Intelligenz (KI) zu fördern und gleichzeitig ein hohes Schutzniveau in Bezug auf Gesundheit, Sicherheit und die in der Charta verankerten Grundrechte, einschließlich Demokratie, Rechtsstaatlichkeit und Umweltschutz, vor schädlichen Auswirkungen von KI-Systemen in der Union zu gewährleisten und die Innovation zu unterstützen.

Mensch

Innovation

Gesundheit, Sicherheit Grundrechte,
Demokratie, Rechtsstaat, Umwelt

FH JOANNEUM
University of Applied Sciences

KI-VO Erwägungsgründe der dazu KI-VO

Gegenstand und Zweck

(1) Zweck dieser Verordnung ist es, das Funktionieren des Binnenmarkts zu verbessern, indem ein einheitlicher Rechtsrahmen insbesondere für die Entwicklung, das Inverkehrbringen, die Inbetriebnahme und die Verwendung von Systemen künstlicher Intelligenz (KI-Systeme) in der Union im Einklang mit den Werten der Union festgelegt wird, um die Einführung von menschenzentrierter und vertrauenswürdiger künstlicher Intelligenz (KI) zu fördern und gleichzeitig ein hohes Schutzniveau in Bezug auf Gesundheit, Sicherheit und der in der Charta der Grundrechte der Europäischen Union („Charta") verankerten Grundrechte, einschließlich Demokratie, Rechtsstaatlichkeit und Umweltschutz, sicherzustellen, den Schutz vor schädlichen Auswirkungen von KI-Systemen in der Union zu gewährleisten und gleichzeitig die Innovation zu unterstützen. Diese Verordnung gewährleistet den grenzüberschreitenden freien Verkehr KI-gestützter Waren und Dienstleistungen, wodurch verhindert wird, dass die Mitgliedstaaten die Entwicklung, Vermarktung und Verwendung von KI-Systemen beschränken, sofern dies nicht ausdrücklich durch diese Verordnung erlaubt wird.

(2) Diese Verordnung sollte im Einklang mit den in der Charta verankerten Werten der Union angewandt werden, den Schutz von natürlichen Personen, Unternehmen, Demokratie und Rechtsstaatlichkeit sowie der Umwelt erleichtern und gleichzeitig Innovation und Beschäftigung fördern und der Union eine Führungsrolle bei der Einführung vertrauenswürdiger KI verschaffen.

Ganze
Wertschöpfungskette

KI-VO Erwägungsgründe der dazu KI-VO

Gegenstand und Zweck

(3) KI-Systeme können problemlos in verschiedenen Bereichen der Wirtschaft und Gesellschaft, auch grenzüberschreitend, eingesetzt werden und in der gesamten Union verkehren. Einige Mitgliedstaaten haben bereits die Verabschiedung nationaler Vorschriften in Erwägung gezogen, damit KI vertrauenswürdig und sicher ist und im Einklang mit den Grundrechten entwickelt und verwendet wird. Unterschiedliche nationale Vorschriften können zu einer Fragmentierung des Binnenmarkts führen und können die Rechtssicherheit für Akteure, die KI-Systeme entwickeln, einführen oder verwenden, beeinträchtigen. Daher sollte in der gesamten Union ein einheitlich hohes Schutzniveau sichergestellt werden, um eine vertrauenswürdige KI zu erreichen, wobei Unterschiede, die den freien Verkehr, Innovationen, den Einsatz und die Verbreitung von KI-Systemen und damit zusammenhängenden Produkten und Dienstleistungen im Binnenmarkt behindern, vermieden werden sollten, indem den Akteuren einheitliche Pflichten auferlegt werden und der gleiche Schutz der zwingenden Gründe des Allgemeininteresses und der Rechte von Personen im gesamten Binnenmarkt auf der Grundlage des Artikels 114 des Vertrags über die Arbeitsweise der Europäischen Union (AEUV) gewährleistet wird. Soweit diese Verordnung konkrete Vorschriften zum Schutz von Einzelpersonen im Hinblick auf die Verarbeitung personenbezogener Daten enthält, mit denen die Verwendung von KI-Systemen zur biometrischen Fernidentifizierung zu Strafverfolgungszwecken, die Verwendung von KI-Systemen für die Risikobewertung natürlicher Personen zu Strafverfolgungszwecken und die Verwendung von KI-Systemen zur biometrischen Kategorisierung zu Strafverfolgungszwecken eingeschränkt wird, ist es angezeigt, diese Verordnung in Bezug auf diese konkreten Vorschriften auf Artikel 16 AEUV zu stützen. Angesichts dieser konkreten Vorschriften und des Rückgriffs auf Artikel 16 AEUV ist es angezeigt, den Europäischen Datenschutzausschuss zu konsultieren.

KI: ja,

aber: harmonisiert und Vorsicht bei Personendaten

KI-VO Erwägungsgründe der dazu KI-VO

Gegenstand und Zweck

(4) KI bezeichnet eine Reihe von Technologien, die sich rasant entwickeln und zu vielfältigem Nutzen für Wirtschaft, Umwelt und Gesellschaft über das gesamte Spektrum industrieller und gesellschaftlicher Tätigkeiten hinweg beitragen. Durch die Verbesserung der Vorhersage, die Optimierung der Abläufe, Ressourcenzuweisung und die Personalisierung digitaler Lösungen, die Einzelpersonen und Organisationen zur Verfügung stehen, kann die Verwendung von KI Unternehmen wesentliche Wettbewerbsvorteile verschaffen und zu guten Ergebnissen für Gesellschaft und Umwelt führen, beispielsweise in den Bereichen Gesundheitsversorgung, Landwirtschaft, Lebensmittelsicherheit, allgemeine und berufliche Bildung, Medien, Sport, Kultur, Infrastrukturmanagement, Energie, Verkehr und Logistik, öffentliche Dienstleistungen, Sicherheit, Justiz, Ressourcen- und Energieeffizienz, Umweltüberwachung, Bewahrung und Wiederherstellung der Biodiversität und der Ökosysteme sowie Klimaschutz und Anpassung an den Klimawandel.

(5) Gleichzeitig kann KI je nach den Umständen ihrer konkreten Anwendung und Nutzung sowie der technologischen Entwicklungsstufe Risiken mit sich bringen und öffentliche Interessen und grundlegende Rechte schädigen, die durch das Unionsrecht geschützt sind. Ein solcher Schaden kann materieller oder immaterieller Art sein, einschließlich physischer, psychischer, gesellschaftlicher oder wirtschaftlicher Schäden.

(6) Angesichts der großen Auswirkungen, die KI auf die Gesellschaft haben kann, und der Notwendigkeit, Vertrauen aufzubauen, ist es von entscheidender Bedeutung, dass KI und ihr Regulierungsrahmen im Einklang mit den in Artikel 2 des Vertrags über die Europäische Union (EUV) verankerten Werten der Union, den in den Verträgen und, nach Artikel 6 EUV, der Charta verankerten Grundrechten und -freiheiten entwickelt werden. Voraussetzung sollte sein, dass KI eine menschenzentrierte Technologie ist. Sie sollte den Menschen als Instrument dienen und letztendlich das menschliche Wohlergehen verbessern.

Mensch

FH JOANNEUM
University of Applied Sciences

Anwendungsbereiche der KI-VO

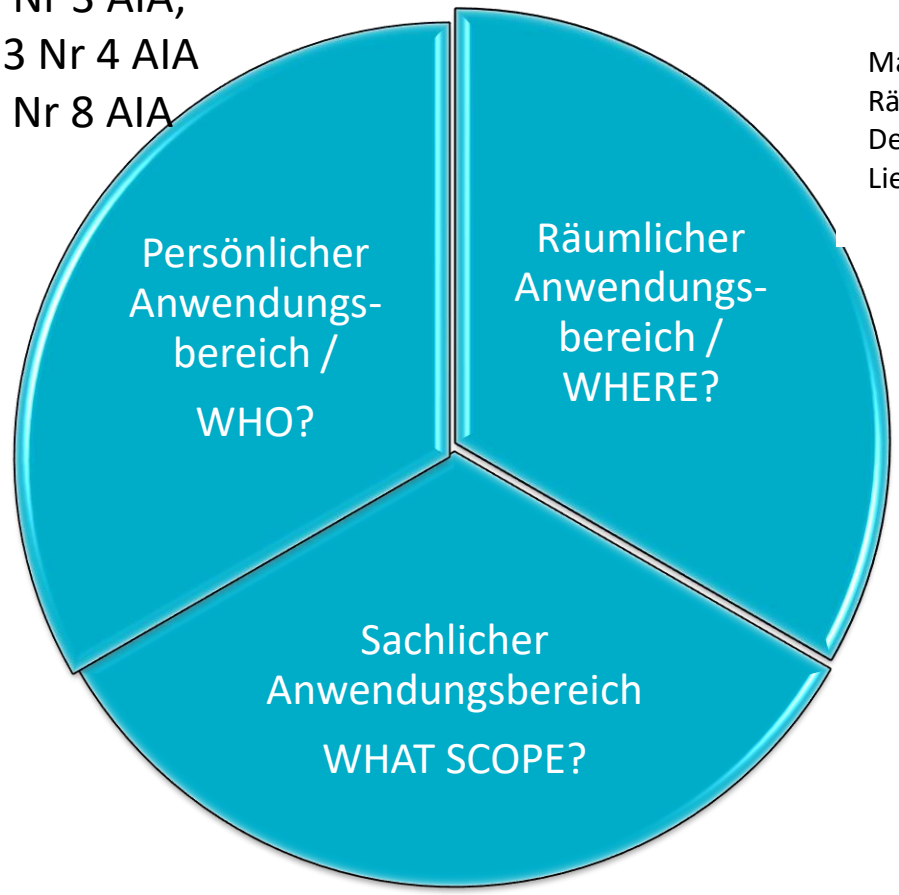
45

Anwendungsbereiche Art 2 AIA und Definitionen Art 3 AIA/ *Scope and Definitions*

„Markt“ Art 2 Abs 1 AIA

Anbieter/*Provider* Art 3 Nr 3 AIA,
Betreiber/*Deployer* Art 3 Nr 4 AIA
Akteure/*Operator* Art 3 Nr 8 AIA
…

Market/Use Art 2 Abs 1 AIA
Räumlich sehr weit, sowohl nach Definition und auch aufgrund der Lieferketten Thematik



Article 25 Responsibilities along the AI value chain

1. Any distributor, importer, deployer or other third-party shall be considered to be a provider of a high-risk AI system for the purposes of this Regulation and shall be subject to the obligations of the provider under Article 16, in any of the following circumstances: …

AI Systeme/*AI Systems* Art 3 Nr 1 AIA, AI Modelle/*AI Models* Art 3 Nr 63 AIA

46

Prüfschritte betreffend die Anwendbarkeit der KI-VO

Für Anwender/Nutzer stellen sich folgende Fragen:

1. Fällt das Produkt oder die Leistung in den **räumlichen Anwendungsbereich** der KI-VO, also wie sieht der Markt aus?
2. Fällt das Produkt oder die Leistung in den **sachlichen Anwendungsbereich** der KI-VO also ist das Gerät / die Software / … ein KI–Service oder ein KI-Modell?
3. Welche Rolle nehme ich ein, muss ich mich an die KI VO halten, bin im **persönlicher Anwendungsbereich,** und wenn ja, dann lautet die Folgefrage:

   in welcher Rolle?

FH JOANNEUM
University of Applied Sciences

KI-Systeme laut Definition Art 3 Nr 1 KI VO

Frage der Anwendbarkeit der KI-VO.

1. Frage lautet: handelt es sich um KI?

FH | JOANNEUM
University of Applied Sciences

Beispiel

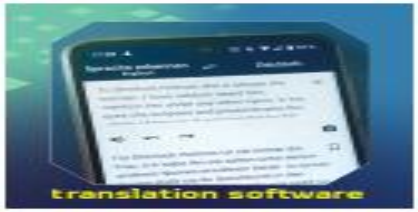
Gehen Sie auf Moodle zum Abschnitt über KI, weiter zum

Arbeitsblatt KI oder keine KI?

Sehen Sie sich die folgenden Beispiele an und entscheiden Sie sich, ob es sich um KI handelt, oder eben nicht.

- *Gruppengröße: ca 3 Personen*
- *Dauer: ca 10 Minuten*

AI SYSTEM



NO AI SYSTEM

FH JOANNEUM
University of Applied Sciences

KI-Systeme laut Definition Art 3 Nr 1 KI VO

Frage der Anwendbarkeit der KI-VO.

1. Frage lautet: handelt es sich um KI? **JA**
2. Handelt es sich um ein KI System oder ein KI Modell?

- Artikel 3: Definition eines KI-Systems als ein vollständig implementiertes System, das mit Menschen oder anderen Systemen interagiert.

- Artikel 10: Anforderungen an die Datenqualität, um Verzerrungen in KI-Modellen zu vermeiden.
- Artikel 13: Transparenzanforderungen für KI-Systeme, damit deren Entscheidungen nachvollziehbar sind.

Die Verordnung konzentriert sich darauf, KI-Systeme sicher, fair und nachvollziehbar zu gestalten. KI-Modelle selbst werden zwar reguliert (z. B. hinsichtlich der Datenqualität), aber die Hauptverantwortung liegt bei den Unternehmen, die komplette KI-Systeme in der Praxis einsetzen.

KI-Modelle mit allgemeiner Verwendbarkeit laut Art 3 Nr 63 KI VO

- Ein KI-Modell ist der mathematische oder algorithmische Kern eines KI-Systems.
- Es handelt sich dabei um eine trainierte statistische Funktion oder ein neuronales Netz, das durch maschinelles Lernen auf Basis von Daten Muster erkennt und Vorhersagen trifft.
- Ein neuronales Netz ist eine Methode des maschinellen Lernens, die auf dem Prinzip der Signalverarbeitung im menschlichen Gehirn basiert. Es besteht aus mehreren Schichten von künstlichen Neuronen, die einfache Berechnungen durchführen und durch das Anpassen von Gewichtungen lernen, Zusammenhänge in den Daten zu erkennen.

KI-Modelle mit allgemeiner Verwendbarkeit laut Art 3 Nr 63 KI VO

- Merkmale eines KI-Modells:

Entwickelt durch maschinelles Lernen (z. B. Entscheidungsbäume oder neuronale Netze).

Funktioniert auf Basis von Trainingsdaten und mathematischen Optimierungen.

Kann Vorhersagen oder Klassifikationen generieren, hat aber keine direkte Interaktion mit Endnutzern.

Ist oft eine „Black Box“. Entscheidungswege sind nicht immer verständlich.

FH JOANNEUM
University of Applied Sciences

KI-Modelle mit allgemeiner Verwendbarkeit laut Art 3 Nr 63 KI VO

- folgende Merkmale:
  - erfüllen ein breites Spektrum unterschiedlicher Aufgaben kompetent;
  - kann in eine Vielzahl nachgelagerter Systeme oder Anwendungen integriert werden
  - ausgenommen KI-Modelle vor ihrem Inverkehrbringen für F&E oder für Konzipierung von Prototypen;

➔ ChatGPD ist KI System

➔ GPT-4 ist KI Modell (KI Programm dahinter)

➔ Beispiele der EU:

KI-Modelle mit allgemeiner Verwendbarkeit laut Art 3 Nr 63 KI VO

Beispiele bekannter KI-Modelle:

GPT-4 ist ein Sprachmodell und wird unter anderem zur Textgenerierung, Übersetzung, für Chatbots und zur Automatisierung von Kundenservice verwendet.

DALL·E 3 ist ein Bildgenerierungsmodell, welches durch die Angabe von Textbeschreibungen Bilder erstellt.

AlphaFold ist ein Deep-Learning-Modell für Proteinfaltung und wird in der Biomedizin verwendet, um 3D-Strukturen von Proteinen vorherzusagen, um neue Medikamente zu entwicken.

FH JOANNEUM
University of Applied Sciences

Rollen in der KI-VO

Kommt man zum Schluss, dass die KI VO räumlich, sachlich und persönlich anwendbar ist und es liegt ein KI-Modell oder KI-System vor, dann lautet die nächste Frage:

Welche Rolle nehme ich als Unternehmen/als Person ein?

Pexels, Angela Roma

Akteure im AI Act

Übersicht: Die Akteure im AI Act

Der AI Act sieht viele Rollen in der KI-Wertschöpfungskette vor. Die verschiedenen Akteure sind dabei keineswegs unbekannt, der Unionsgesetzgeber orientierte sich dabei in vielerlei Hinsicht an den EU-Produktrechtsnormen (siehe etwa die Produktsicherheitsverordnung, Medizinprodukteverordnung).

Zu den Akteuren im AI Act zählen gemäß Art. 3 Ziffer 8:

- **Anbieter („Provider");**
- **Produkthersteller („Product Manufacturer");**
- **Bevollmächtigter („Authorised Representative");**
- **Einführer („Importer");**
- **Händler („Distributor");**
- **Betreiber („Deployer").**

Ferner kommen auch noch Nutzer und „betroffene Personen" vor. Diese werden aber nicht als Akteure im Sinne des AI Act bezeichnet.

FH JOANNEUM
University of Applied Sciences

Wer ist betroffen?

Die KI-Verordnung betrifft eine breite Palette von Akteuren, die mit der Entwicklung, dem Verkauf, der Bereitstellung oder der Nutzung von KI-Systemen in der Europäischen Union befasst sind. Konkret betrifft die Verordnung:

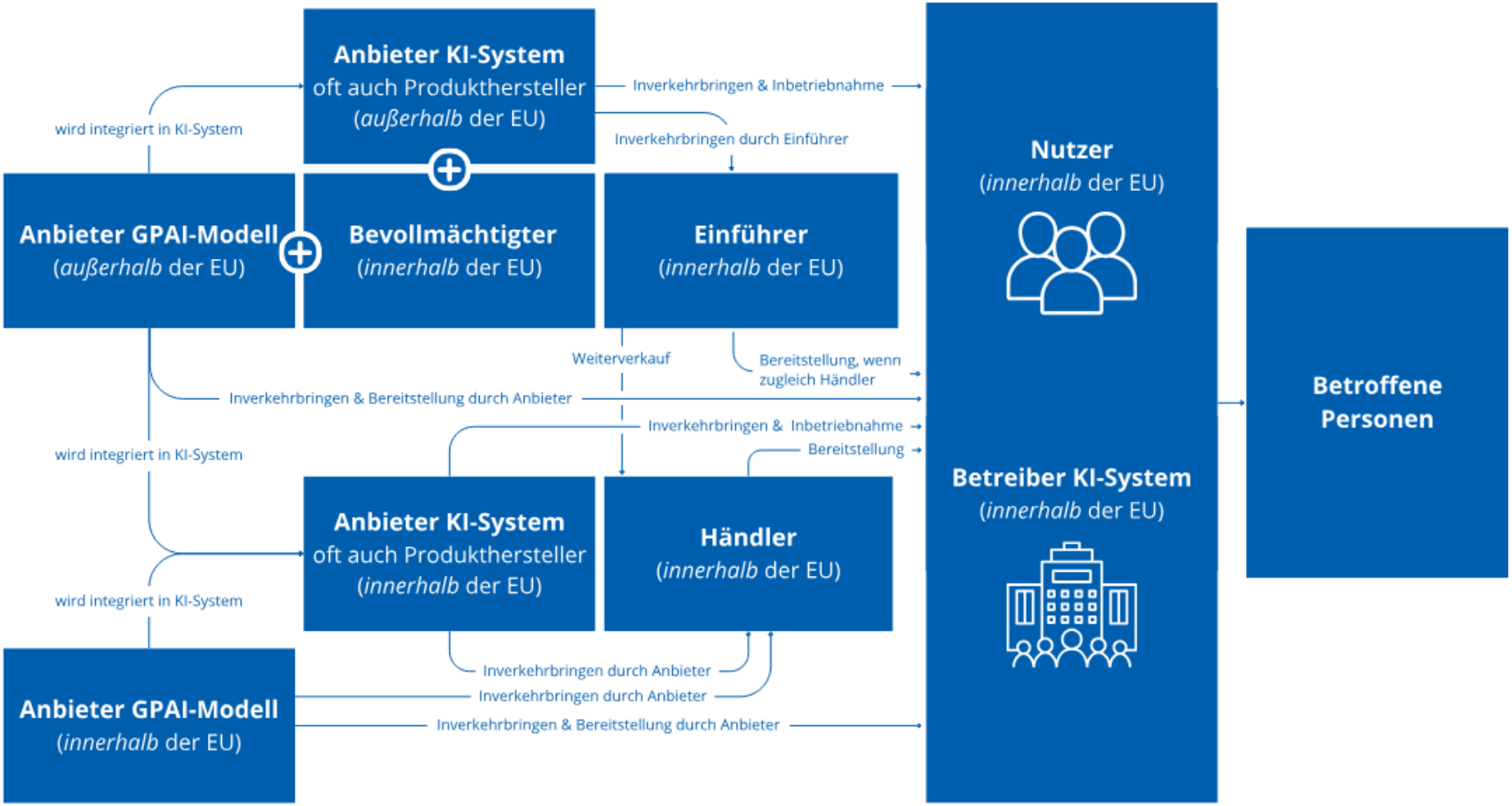
1. Anbieter von KI-Systemen: Unternehmen und Einzelpersonen, die KI-Systeme entwickeln oder herstellen, sind für die Einhaltung der Verordnung verantwortlich. Dies umfasst die Durchführung von Risikobewertungen, die Einhaltung von Transparenz- und Dokumentationsanforderungen, die Sicherstellung der Datenqualität und die Kennzeichnung von KI-Systemen. Sie müssen zudem Konformitätsbewertungen durchführen, um zu gewährleisten, dass ihre Systeme den gesetzlichen Anforderungen entsprechen.
2. 2. Nutzer von KI-Systemen: Organisationen und Einzelpersonen, die KI-Systeme verwenden, müssen sicherstellen, dass diese Systeme im Einklang mit der Verordnung eingesetzt werden. Dies schließt die ordnungsgemäße Verwendung, Überwachung und gegebenenfalls Anpassung von KI-Systemen ein, um den gesetzlichen Anforderungen gerecht zu werden.

Akteure im AI Act

3. Importeure und Distributoren: Akteure, die KI-Systeme in die EU einführen oder innerhalb der EU vertreiben, müssen überprüfen, dass die Produkte den Anforderungen der Verordnung entsprechen. Sie sind verantwortlich dafür, die notwendige Dokumentation zu beschaffen und sicherzustellen, dass die Systeme korrekt gekennzeichnet sind.
4. Dritte, die KI-Systeme modifizieren: Personen oder Organisationen, die bestehende KI-Systeme erheblich verändern, müssen sicherstellen, dass diese modifizierten Systeme den gesetzlichen Anforderungen entsprechen. Dies kann zusätzliche Konformitätsbewertungen und die Anpassung der Dokumentation erforderlich machen.
5. Nationale Aufsichtsbehörden und das europäische KI-Gremium: Diese Institutionen sind für die Überwachung der Einhaltung der Verordnung verantwortlich. Sie führen Inspektionen durch, bearbeiten Beschwerden, bewerten die Konformität von KI-Systemen und setzen Sanktionen bei Verstößen durch.
6. Dienstleister für Daten und Datenverarbeitung: Unternehmen, die Daten für das Training oder den Betrieb von KI-Systemen bereitstellen oder verarbeiten, müssen sicherstellen, dass diese Daten den Anforderungen an Qualität, Repräsentativität und Datenschutz entsprechen. Sie tragen zur Einhaltung der Verordnung bei, indem sie die Voraussetzungen für den rechtskonformen Einsatz von KI schaffen

AI Act: Akteure

Die Rollen entlang der KI-Wertschöpfungskette



FH JOANNEUM
University of Applied Sciences

Anbieter und Betreiber als Hauptadressaten Art 3 Nr 3 und 4 KI-VO

- Anbieter:
  - eine natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System oder ein KI-Modell mit allgemeinem Verwendungszweck entwickelt oder entwickeln lässt und es unter ihrem eigenen Namen oder ihrer Handelsmarke in Verkehr bringt oder das KI-System unter ihrem eigenen Namen oder ihrer Handelsmarke in Betrieb nimmt, sei es entgeltlich oder unentgeltlich;
- Betreiber:
  - natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System in eigener Verantwortung verwendet,

➔ uU auch weitere Akteure wie Händler, Importeur ….

Doppelrollen

- Durch die Regel in Art 25 Abs 1 AIA ist es auch möglich, dass Einführer, Händler, Betreiber und sonstige Dritte als Anbieter behandelt werden, also eine Doppelrolle einnehmen, wenn sie:

- ein bereits in Verkehr gebrachtes oder in Betrieb genommenes Hochrisiko-KI-System mit ihrem Namen oder ihrem Unternehmenskennzeichen versehen wird , unbeschadet vertraglicher Vereinbarungen, die eine andere Aufteilung der Pflichten vorsieht. Der jeweilige Akteur tritt somit als Anbieter eines KI-Systems auf, obwohl dieses nicht selbst entwickelt wurde („Quasi-Hersteller“).

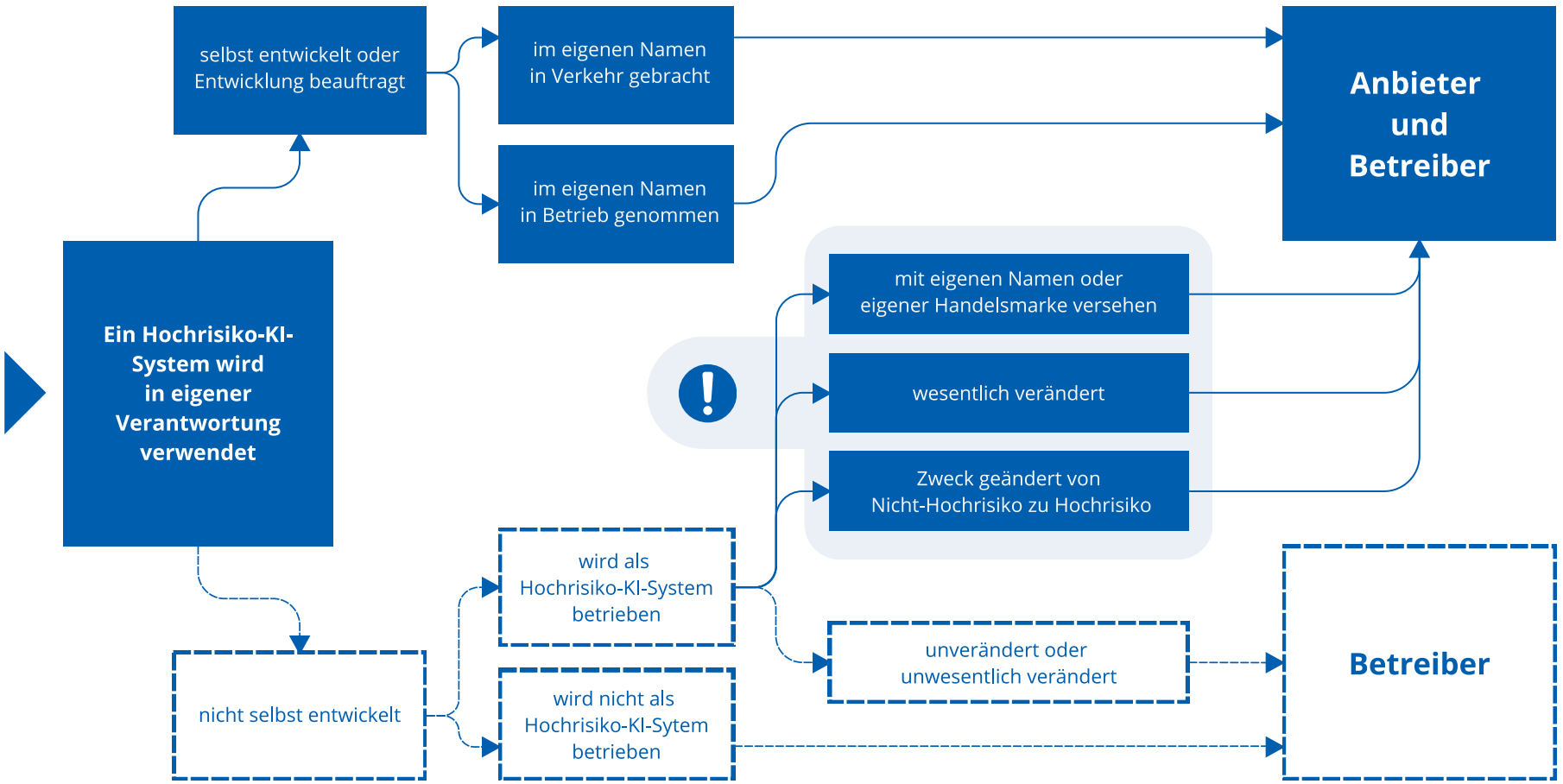
- eine wesentliche Änderung an einem Hochrisiko-KI-System, das bereits in Verkehr gebracht oder in Betrieb genommen wurde, so vornehmen, dass es weiterhin ein Hochrisiko-KI-System im Sinne von Art 6 AIA bleibt. Eine wesentliche Änderung liegt vor, wenn diese ein KI-System nach Inverkehrbringen oder Inbetriebnahme, so verändert, dass dies nicht in der ursprünglichen Konformitätsbewertung vorgesehen oder geplant war und dadurch die Konformität des KI-Systems beeinträchtigt wird oder die zu einer Änderung der Zweckbestimmung führt, für die das KI-System bewertet wurde, Art 3 Z 23 AI Act.

- die Zweckbestimmung eines KI-Systems, einschließlich eines GPAI-Systems, das nicht als hochriskant eingestuft wurde und bereits in Verkehr gebracht oder in Betrieb genommen wurde, so verändern, dass das betreffende KI-System zu einem Hochrisiko-KI-System im Sinne von Art 6 AIA wird.

Zusammengefasst: Sobald ein Akteur ein KI-System mit seinem Namen versieht, es stark verändert oder die Zweckbestimmung ändert, nachdem es bereits in Verkehr gebracht oder in Betrieb genommen wurde, wird er als Quasi-Hersteller und somit als Anbieter mit entsprechenden Pflichten gesehen.

Rollenverteilung im Hochrisiko-Bereich

Je nach Einsatzart eines Hochrisiko-KI-Systems gelten Anbieter- oder Betreiberpflichten



FH JOANNEUM
University of Applied Sciences

Rollen, wer fehlt?

Legal

Provider/Anbieter Art 3 Nr 3 AIA
Deployer/Betreiber Art 3 Nr 4 AIA
Operator/Akteur Art 3 Nr 8 AIA
…

Market principle (partly)/ „Marktortprinzip“ Art 2 Abs 1 AIA mit Ausnahmen

AI Systems Art 3 Nr 1
General purpose AI models Art 3 Nr 63 AIA

Users/Nutzer?
Operators?
In seiner Verwantwortung ….

Affected Persons/Betroffene Person?

Developer/Entwickler:in?
Dh Entwickler:in programmiert Recht, ohne das zu überdenken, Bsp Tesla Auto ohne Sensor unter dem Auto.

Risikostufen der KI-VO

Kommt man zum Schluss, die KI-VO ist räumlich, sachlich und persönlich zur Anwendung zu bringen, dann lautet die nächste Frage:

Welches Risiko bringt die KI bzw deren Anwendung mit sich, also in welcher Risikostufe ist meine Anwendung?

FH JOANNEUM
University of Applied Sciences

Risiko Pyramide(n)

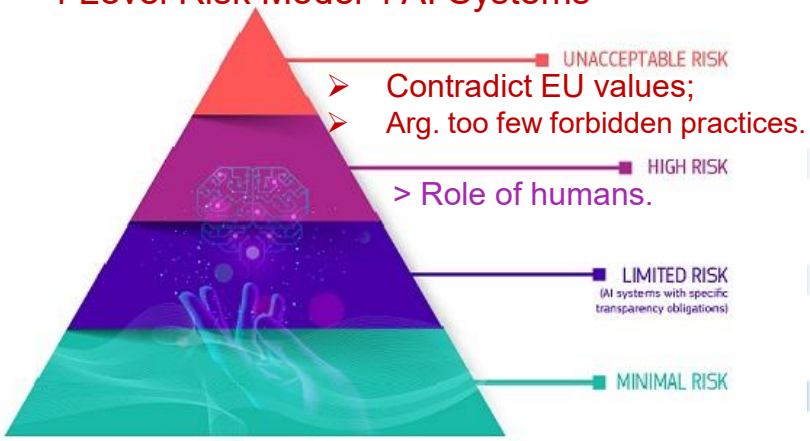
- 1. Schritt: Anwendbarkeit der KI-VO ✓
- 2. Rolle(n) ✓
- 3. KI-System oder KI-Modell ✓
- 4. Schritt: welche Stufe der Pyramide(n) ✓
- Pflichten:
  - Transparenz
  - KI-Kompetenz
  - Menschliche Aufsicht
  - …

FH JOANNEUM
University of Applied Sciences

The human in the AI Act – Risk pyramide(s)

Legal

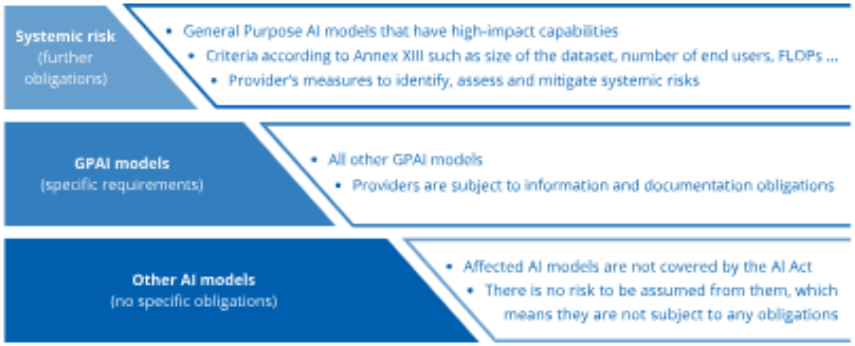
4 Level Risk Model 4 AI Systems



- Contradict EU values;
- Arg. too few forbidden practices.

> Role of humans.

3 Level Risk Model 4 AI Models

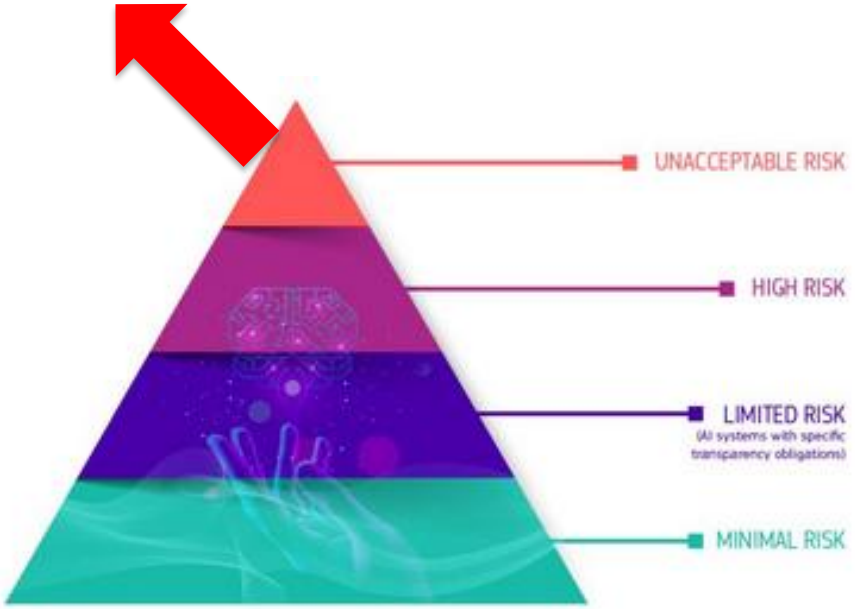


® RTR Austrian Regulatory Authority for Broadcasting and Telecommunications

FH JOANNEUM
University of Applied Sciences

Risikopyramide für AI Systeme – Verbote Art 5 Abs 1 lit a –h AIA

- Alles, was als eindeutige Bedrohung für EU-Bürger angesehen wird, wird verboten: von der behördlichen Bewertung des sozialen Verhaltens
  - Social Scoring



FH JOANNEUM
University of Applied Sciences

Verbotene Praktiken ab 2.2.2025

- KI-Systeme, die Personen beeinflussen, indem **unterschwellige** Techniken zur Beeinflussung **oder absichtlich manipulative oder täuschende Techniken** eingesetzt werden, mit dem Ziel oder der Wirkung, das **Verhalten** einer Person oder einer Gruppe **wesentlich zu verändern** und dabei ihren freien Willen zu umgehen, wodurch ihr ein **erheblicher Schaden** zugefügt wird oder werden kann;

- KI-Systeme, die die **Schwächen von Menschen** (zB aufgrund ihres Alters, ihrer Behinderung oder sozialen oder wirtschaftlichen Situation) ausnutzen, wodurch diesen ein **erheblicher Schaden** zugefügt wird oder werden kann;

- KI-Systeme, die Personen oder Gruppen auf Grundlage ihres sozialen Verhaltens oder persönlicher Eigenschaften oder Merkmale bewerten oder klassifizieren, wobei die dadurch hergeleitete soziale Bewertung zu einer Schlechterstellung oder Benachteiligung dieser Personen oder Gruppen führt (**sog Social Scoring**);

- KI-Systeme, die gezielt Gesichtsbilder aus dem Internet oder Überwachungsaufnahmen auslesen, um **Gesichtserkennungsdatenbanken zu erstellen**;

- KI-Systeme, die **Emotionserkennung von Personen am Arbeitsplatz oder in Bildungseinrichtungen** durchführen (mit Ausnahmen im medizinischen Bereich oder aus Sicherheitsgründen zB Müdigkeitserkennung von Piloten oder Berufskraftfahrern;

- KI-Systeme, **die biometrische Kategorisierungen** (für Art 9 DSGVO Infos) von Personen durchführen, um deren Rasse, politische Einstellung, Gewerkschaftszugehörigkeit, religiöse oder weltanschauliche Überzeugung, Sexualleben oder sexuelle Ausrichtung daraus abzuleiten (mit Ausnahme im Bereich der Strafverfolgung).

Verbotene KI-Systeme im öffentlichen Bereich:

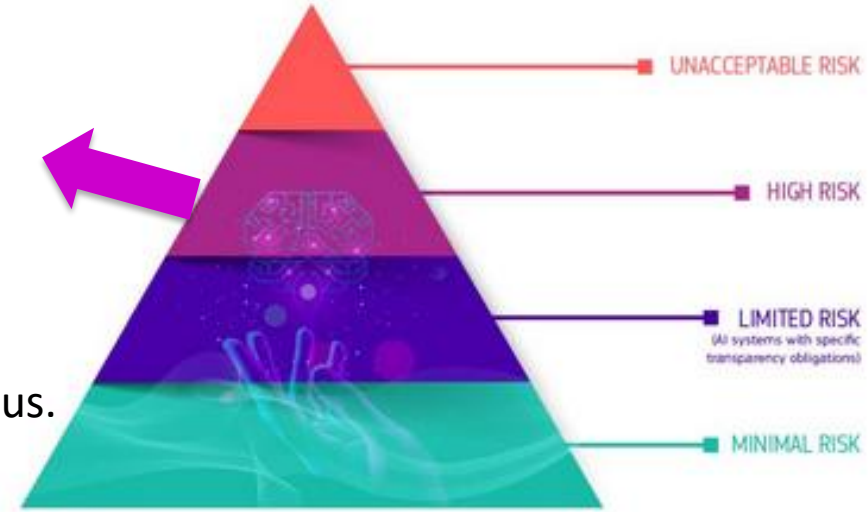
- KI-Systeme, die **Risikobewertungen auf Grundlage von Profiling oder Bewertung persönlicher Merkmale** durchführen und damit eine Risikobewertung vornehmen, ob eine natürliche Person **künftig eine Straftat** begehen wird;

- KI-Systeme, die **die biometrische Echtzeit-Fernerkennung in öffentlich zugänglichen Räumen** zu Strafverfolgungszwecken durchführen (mit Ausnahmen, z.B. bei vermissten Kindern, Tätern bei bestimmten Straftaten).

Risikopyramide für AI Systeme – Hochrisiko Art 6 bis 49

- Kritische Infrastrukturen zB Verkehr, Leben und die Gesundheit
- Zugang zu Schul- oder Berufsausbildung, z B Bewertung von Prüfungen
- Sicherheitskomponenten von Produkten zB KI-Anwendung für die roboterassistierte Chirurgie
- Beschäftigung, Personalmanagement und Zugang zu selbstständiger Tätigkeit zB Software zur Auswertung von Lebensläufen für Einstellungsverfahren
- Zentrale private und öffentliche Dienstleistungen zB Bewertung der Kreditwürdigkeit
- Strafverfolgung zB Überprüfung der Echtheit von Beweismitteln
- Migration, Asyl und Grenzkontrolle
  zB Überprüfung Reisedokumente
- Rechtspflege und demokratische Prozesse

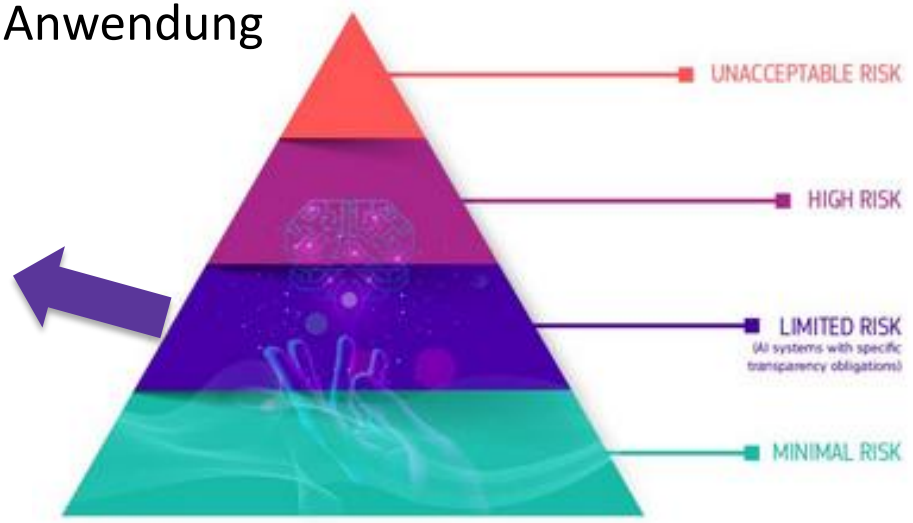
Alle Systeme werden sorgfältig geprüft,
bevor sie in Verkehr gebracht werden
und auch während ihres gesamten Lebenszyklus.



FH JOANNEUM
University of Applied Sciences

Risikopyramide für AI Systeme – begrenztes Risiko Art 50 AIA

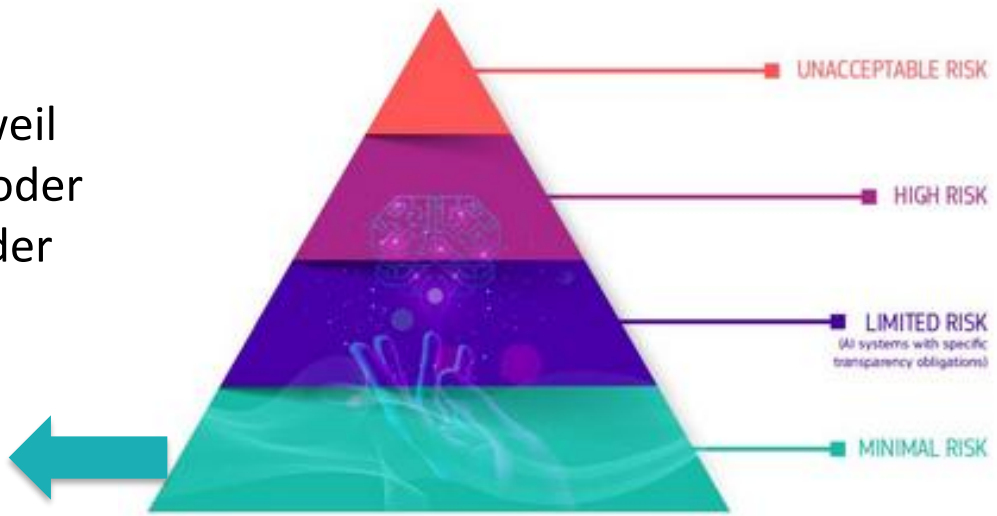
- Für KI-Systeme wie „Chatbots“ gelten minimale Transparenzverpflichtungen, die es den mit ihnen interagierenden Nutzern ermöglichen sollen, fundierte Entscheidungen zu treffen. Die Nutzer können dann entscheiden, ob sie die Anwendung weiter nutzen oder nicht.



FH JOANNEUM
University of Applied Sciences

Risikopyramide für AI Systeme – Minimales Risiko

- Kostenlose Nutzung von Anwendungen wie KI-gestützten Videospielen oder Spamfiltern. Unter diese Kategorie, in der die neuen Vorschriften nicht greifen, fällt die große Mehrzahl der KI-Systeme, weil diese Systeme nur ein minimales oder kein Risiko für die Bürgerrechte oder die Sicherheit darstellen.



FH JOANNEUM
University of Applied Sciences

KI Modelle

= KI

Bsp: ChatGPD = KI System

GPD 3 = Modell

2 Risikostufen Art 51 AIA: mit systemischen Risiko und ohne

AI Act: Risikostufen für KI-Systeme

Nicht alle KI-Systeme fallen in den regulierten Bereich - je höher das Risiko, desto strikter die Regeln

Inakzeptables Risiko
(verboten)

- Verboten, weil sie im Widerspruch zu den Werten der EU stehen
- z. B. KI-Systeme, die das menschliche Verhalten manipulieren oder Schwächen ausnutzen oder auch "Social Scoring" und "Predictive Policing"

Hohes Risiko
(Konformitätsbewertung)

- Anforderungen an Inverkehrbringen bzw. Inbetriebnahme
- KI-Systeme in bestimmten Produkten und Bereichen (Anhang I & III)
- z. B. Spielzeug, Zivilluftfahrt, Biometrik, kritische Infrastruktur

Begrenztes Risiko
(Transparenzverpflichtung)

- Risikomanagement durch Transparenzmaßnahmen
- z. B. Chatbots oder KI-Systeme zur Erstellung von Text, Audio, Bild oder Video

Minimales bzw. kein Risiko
(keine spezifischen Verpflichtungen)

- Alle anderen KI-Systeme
- z. B. Videospiele, Spam-Filter
- Freiwillige Verhaltenskodizes

AI Act: Risikostufen für KI-Modelle

Nicht alle KI-Modelle fallen in den regulierten Bereich - je höher das Risiko, desto strikter die Regeln

Systemisches Risiko
(weitergehende Pflichten)

- General-Purpose-AI-Modelle, die Fähigkeiten mit hohem Wirkungsgrad besitzen
- Kriterien gemäß Anhang XIII wie z. B. Größe des Datensatzes, Endnutzerzahl, FLOPs etc.
- Anbieter-Maßnahmen zur Ermittlung, Bewertung und Minderung von Systemrisiken

GPAI-Modelle
(spezifische Anforderungen)

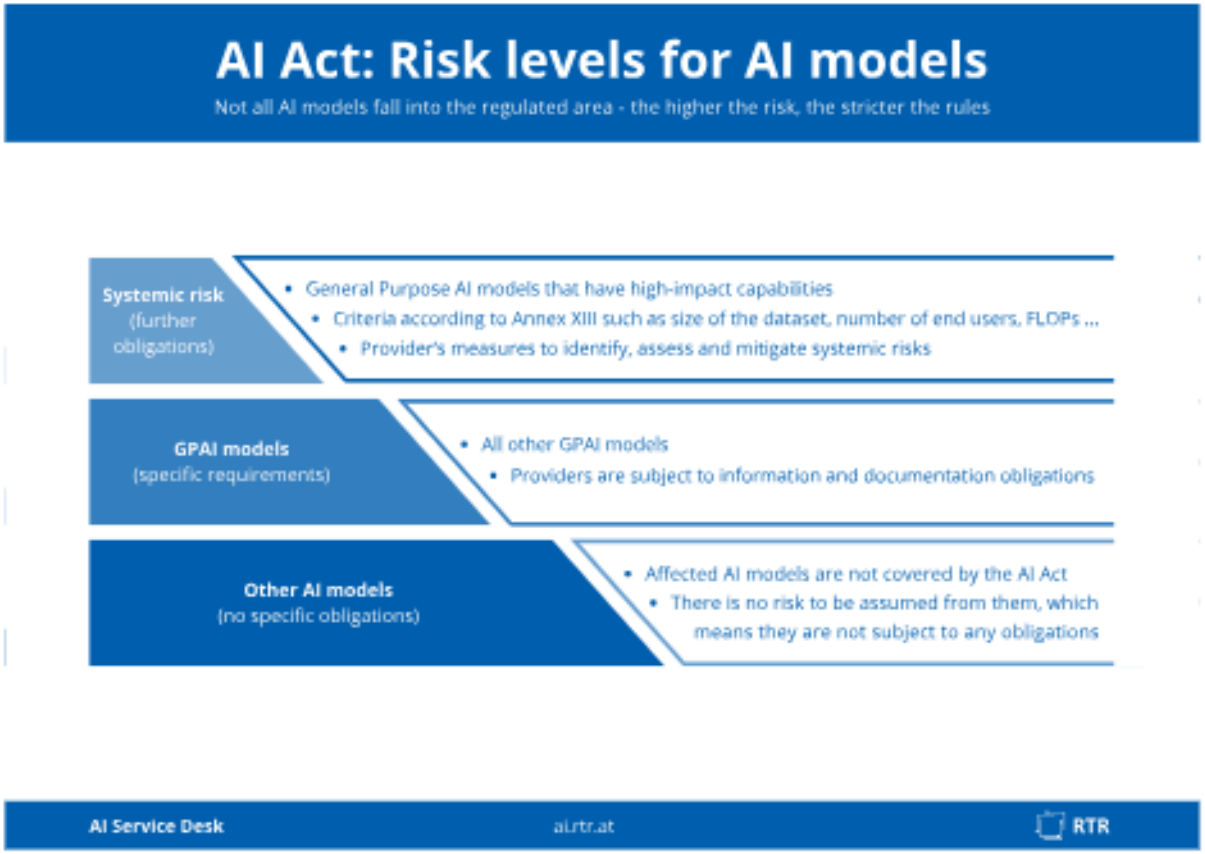
- Alle anderen GPAI-Modelle
- Anbieter unterliegen Informations- und Dokumentationspflichten

Sonstige KI-Modelle
(keine spezifischen Verpflichtungen)

- Betroffene KI-Modelle sind nicht vom AI Act erfasst
- Von ihnen ist kein Risiko anzunehmen, wodurch sie keine Verpflichtungen auferlegt bekommen

FH JOANNEUM
University of Applied Sciences

KI Modelle



GPAI models are AI models that are capable of handling a wide range of tasks. Certain obligations apply to them. In the event of a systemic risk, the catalogue of obligations expands © RTR (CC BY 4.0)

FH JOANNEUM
University of Applied Sciences

Pflichten der KI-VO

Die letzte Frage lautet, welche Pflichten treffen mich in meiner Rolle mit der KI in der jeweiligen Risikostufe.

Transparenz

Kompetenz

Aufsicht

AI Act: Verpflichtungen von Betreibern

Der Umfang der Verpflichtungen nimmt entsprechend der Risikoklassifizierung des KI-Systems ab

	Hochrisiko KI-System	**KI-System** begrenzt. Risiko	**KI-System** minimal. Risiko
KI-Kompetenz	Art. 4	Art. 4	Art. 4
Transparenz gegenüber nachgelagerten Akteuren	Art. 26 (11)	Art. 50 (3), (4)	
Verwendung des KI-Systems laut Betriebsanleitung	Art. 26 (1), (3), (4)		
Menschliche Aufsicht	Art. 26 (2)		
Überwachung des KI-Systems	Art. 26 (5)		
Meldung von schwerwiegenden Vorfällen	Art. 26 (5), 73		
Aufbewahrung von erzeugten Protokollen	Art. 26 (6)		
Sofern relevant, Datenschutz-Folgenabschätzung	Art. 26 (9)		
Zusammenarbeit mit zuständigen nationalen Behörden	Art. 26 (12)		
Recht auf Erläuterung der Entscheidungsfindung im Einzelfall	Art. 86 (1)		
Informationspflichten gegenüber der Arbeitnehmer:innen-Vertretung *sofern Arbeitgeber:in Hochrisiko-KI-Systeme am Arbeitsplatz einsetzt*	Art. 26 (7)		
Registrierungspflicht *sofern EU-Organe, EU-Einrichtungen und sonstige EU-Stellen*	Art. 26 (8), 49		
Genehmigungspflicht einer Justiz- oder Verwaltungsbehörde *sofern Einsatz zur nachträglichen biometrischen Fernidentifizierung*	Art. 26 (10)		
Erstellung einer Grundrechte-Folgenabschätzung *sofern u. a. öffentl. oder private Einrichtungen öffentliche Dienste erbringen*	Art. 27		

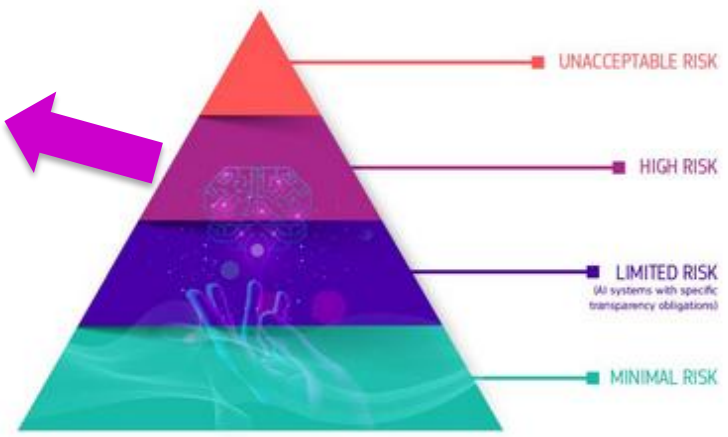
AI Act: Verpflichtungen von Anbietern

Der Umfang der Verpflichtungen nimmt entsprechend der Risikoklassifizierung des KI-Systems/KI-Modells ab

	Hochrisiko KI-System	**GPAI-Modell** system. Risiko	**GPAI-Modell**	**KI-System** begrenzt. Risiko	**KI-System** minimal. Risiko
KI-Kompetenz	Art. 4	Art. 4	Art. 4	Art. 4	Art. 4
Transparenz gegenüber nachgelagerten Akteuren	Art. 13	Art. 55 (1)	Art. 53 (1) b	Art. 50 (1), (2)	
Anforderungen an Daten	Art. 10	Art. 55 (1)	Art. 53 (1) c, d		
Technische Dokumentation	Art. 11	Art. 55 (1)	Art. 53 (1) a		
Zusammenarbeit mit Behörden	Art. 21	Art. 55 (1)	Art. 53 (3)		
Bennenung Bevöllmächtigter (sofern Drittstaat)	Art. 22	Art. 55 (1)	Art. 54		
Risikomanagement	Art. 9	Art. 55 (1) a, b			
Genauigkeit, Robustheit und Cybersicherheit	Art. 15	Art. 55 (1) d			
Registrierungs- bzw. Mitteilungspflichten	Art. 49	Art. 52 (1)			
Meldepflichten gegenüber Behörden	Art. 73	Art. 55 (1) c			
Aufzeichnung von Ereignissen	Art. 12				
Implementierung menschlicher Überwachungstools	Art. 14				
Kennzeichnungspflichten	Art. 16 b				
Sicherstellung der Barrierefreiheitsanforderungen	Art. 16 l				
Qualitätsmanagement	Art. 17				
Aufbewahrungspflichten	Art. 18, 19				
Korrekturmaßnahmen	Art. 20				
Konformitäts-Bewertungsverf., -Erklärung, -Kennzeichnung	Art. 43, 47, 48				

Pflichten für hochrisiko KI-Systeme – Art 6 bis 49 KI.VO

- Stand der Technik beachten
- Risikomanagementsystem / Qualitätsmanagementsystem (Art. 9, 17 KI-VO) muss vorhanden sein
- Resilienz / Cyber-Security (Art. 15 KI-VO)
- Testverfahren / regelmäßige Tests (Art. 16 KI-VO) müssen stattfinden
- EU-Konformitätserklärung und CE-Kennzeichnung (Art. 16 KI-VO) müssen vorliegen
- Zusammenarbeit mit den zuständigen Behörden (Art. 21 KI-VO)
- Korrekturmaßnahmen / Informationspflichten müssen umgesetzt werden (Art. 20 KI-VO)
- technische Dokumentation / Aufbewahrungspflichten (Art. 11, 18 KI-VO)
- Transparenz (Art. 13 KI-VO)
- Protokollierung von Funktionsmerkmalen (Art. 12 KI-VO)
- Beobachtung nach Markteinführung (Art. 71 KI-VO)
- Meldung von „schwerwiegenden Vorfällen“ (Art. 73 KI-VO)
- Durchführung einer Grundrechte-Folgenabschätzung (Art. 27 KI-VO)
- Entwicklung mit Trainingsdaten, die eine bestimmte Qualität aufweisen (s genannte Daten-Governance, Art. 10 KI-VO)
- Registrierung (Art. 49 KI-VO)
- menschliche Aufsicht (Art. 14 KI-VO)
- Maßnahmen zur Barrierefreiheit (Art. 16 KI-VO)



Der verpflichtete Akteur muss diese Pflichten nachweisen. Primär die KI-Anbieter, also die KI-Hersteller, in manchen Fällen aber auch die Betreiber. Darunter wird jede natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle verstanden, die ein KI-System in eigener Verantwortung verwendet, es sei denn, das KI-System wird im Rahmen einer persönlichen und nicht beruflichen Tätigkeit verwendet (Art. 3 Nr. 4 KI-VO).

Pflichten der KI-VO

- Kennzeichnungspflicht: KI-Systeme müssen klar als solche gekennzeichnet sein.
- Dokumentation und Transparenz: Insbesondere für hochriskante Systeme müssen Unternehmen umfassende Dokumentation und Transparenz sicherstellen.
- Datenmanagement: Sicherstellung der Qualität und Repräsentativität der verwendeten Daten.
- Zusammenarbeit mit Aufsichtsbehörden: Bereitschaft zur Offenlegung und Zusammenarbeit bei Audits und Kontrollen.

Anbieter und Betreiber als Hauptadressaten 4 AIA

Artikel 4 AI Act verpflichtet Anbieter und Betreiber, intern KI-Kompetenz zu schaffen. Betreiber sind all jene, die KI in ihrer Organisation bloß anwenden, egal welchen Risikograd die KI hat, dh die Verpflichtung gilt auch beim Einsatz von KI mit bloß minimalem Risiko.

Erwägungsgrund 20

Allgemeine Pflicht im Art 4 AIA (KI-Kompetenz/AI Literacy)

„Die Anbieter und Betreiber von KI-Systemen ergreifen Maßnahmen, um nach besten Kräften sicherzustellen, dass ihr Personal und andere Personen, die in ihrem Auftrag mit dem Betrieb und der Nutzung von KI-Systemen befasst sind, über ein ausreichendes Maß an KI-Kompetenz verfügen, wobei ihre technischen Kenntnisse, ihre Erfahrung, ihre Ausbildung und Schulung und der Kontext, in dem die KI-Systeme eingesetzt werden sollen, sowie die Personen oder Personengruppen, bei denen die KI-Systeme eingesetzt werden sollen, zu berücksichtigen sind.“

✓ Training!

Anbieter und Betreiber als Hauptadressaten 4 AIA

Anbieter und Betreiber müssen sicherstellen, dass ihr Personal und jenes ihrer Auftragsverarbeiter, die KI-Systeme betreiben, über ein ausreichendes Maß an Kompetenz im Bereich KI verfügen, wobei diesbezüglich deren

- technische Kenntnisse,
- Erfahrung,
- Ausbildung und Schulung zu berücksichtigen sind.

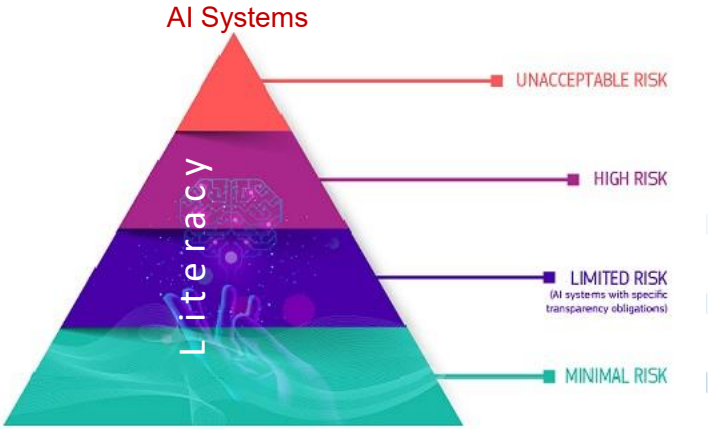
Weiters sind dabei zu beachten

- der Kontext, in dem die KI-System eingesetzt werden sollen und
- die Personen / -gruppen, bei denen die KI-Systeme eingesetzt werden sollen.

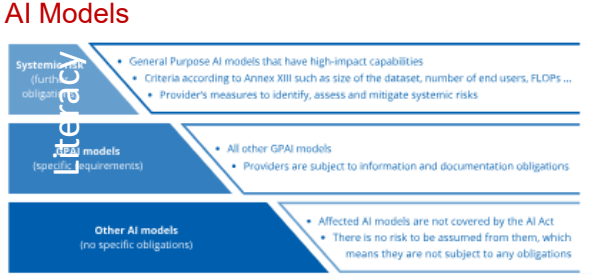
In der Praxis bedeutet dies, dass Mitarbeiterinnen und Mitarbeiter, die noch keine ausreichenden technischen Kenntnisse und Erfahrungen haben, für die KI-Anwendung ausgebildet oder geschult werden müssen.

FH JOANNEUM
University of Applied Sciences

KI-Kompetenz (AI-Literacy) Art 4 KI-VO (gem Def° Art 3 Nr 56 KI-VO)



Reasoning 20: In order to obtain the greatest benefits from AI systems … AI literacy should equip providers, deployers and affected persons … to make informed decisions regarding AI systems.



® RTR Austrian Regulatory Authority for Broadcasting and Telecommunications

Literacy affects primarily **providers**, **deployers** as an obligation for AI systems, also relevant for **processors** (who operate AI systems) and **affected persons**.

Branley-Bell / Proßnegg / Feiner / Libal, IRIS 2025, Vienna

FH JOANNEUM
University of Applied Sciences

KI-Kompetenz Kompetenz Erwägungsgrund 20 KI-VO

Um den größtmöglichen Nutzen aus KI-Systemen zu ziehen und gleichzeitig die Grundrechte, Gesundheit und Sicherheit zu wahren und eine demokratische Kontrolle zu ermöglichen, sollte die KI-Kompetenz Anbieter, Betreiber und betroffene Personen mit den notwendigen Konzepten ausstatten, um fundierte Entscheidungen über KI-Systeme zu treffen. Diese Konzepte können in Bezug auf den jeweiligen Kontext unterschiedlich sein und das Verstehen der korrekten Anwendung technischer Elemente in der Entwicklungsphase des KI-Systems, der bei seiner Verwendung anzuwendenden Maßnahmen und der geeigneten Auslegung der Ausgaben des KI-Systems umfassen sowie — im Falle betroffener Personen — das nötige Wissen, um zu verstehen, wie sich mithilfe von KI getroffene Entscheidungen auf sie auswirken werden. Im Zusammenhang mit der Anwendung dieser Verordnung sollte die KI-Kompetenz allen einschlägigen Akteuren der KI-Wertschöpfungskette die Kenntnisse vermitteln, die erforderlich sind, um die angemessene Einhaltung und die ordnungsgemäße Durchsetzung der Verordnung sicherzustellen. Darüber hinaus könnten die umfassende Umsetzung von KI-Kompetenzmaßnahmen und die Einführung geeigneter Folgemaßnahmen dazu beitragen, die Arbeitsbedingungen zu verbessern und letztlich die Konsolidierung und den Innovationspfad vertrauenswürdiger KI in der Union unterstützen. Ein Europäisches Gremium für Künstliche Intelligenz (im Folgenden „KI-Gremium“) sollte die Kommission dabei unterstützen, KI-Kompetenzinstrumente sowie die Sensibilisierung und Aufklärung der Öffentlichkeit in Bezug auf die Vorteile, Risiken, Schutzmaßnahmen, Rechte und Pflichten im Zusammenhang mit der Nutzung von KI-Systeme zu fördern. In Zusammenarbeit mit den einschlägigen Interessenträgern sollten die Kommission und die Mitgliedstaaten die Ausarbeitung freiwilliger Verhaltenskodizes erleichtern, um die KI-Kompetenz von Personen, die mit der Entwicklung, dem Betrieb und der Verwendung von KI befasst sind, zu fördern.

AI Literacy Art 4 AI Act

Pexels.com: Vlada Karpovich.

✓ Training of staff and processors

? affected persons, developers?

☑ I confirm my AI competence.

The context in which the AI systems are to be used:

The persons/groups for whom the AI systems are to be used.

Example for AI bias.

Branley-Bell / Proßnegg / Feiner / Libal, IRIS 2025, Vienna

Literacy

Technology

Artikel 4 gilt ungeachtet der Risikoeinstufung der KI-Anwendung und verpflichtet Anbieter bzw. Betreiber von KI-Systemen, ein ausreichendes Maß an „KI-Kompetenz“ im Unternehmen sicherzustellen. Die Erläuterungen spezifizieren, dass Mitarbeiter*innen und andere Personen mit den notwendigen Kompetenzen auszustatten sind, um fundierte Entscheidungen über KI-Systeme treffen und die angemessene Einhaltung der rechtlichen Vorschriften sicherstellen zu können. Zentral ist außerdem, dass jene Personen, die mit dem KI-System operieren, dessen Risiken und Chancen verstehen, um das System sachkundig einzusetzen und sich möglicher Schäden bewusst zu werden.

Dies beinhaltet im Konkreten folgende Kompetenzen:

- **Risikoabschätzung**: Welche Risiken bestehen beim Einsatz des infragestehenden Systems, insbesondere in Bezug auf die Gesundheit, Sicherheit und Grundrechte, und welche Auswirkungen haben KI-basierte Entscheidungen auf betroffene Personen?
- **Rechtliche Kenntnisse**: Welche rechtlichen Anforderungen sind einzuhalten, um Compliance mit dem AI Act zu gewährleisten?
- **Technische Kenntnisse**: Wie können die technischen Elemente des Systems korrekt angewendet bzw. dessen Ausgaben geeignet interpretiert werden?
- **Unterschiedliche Ausgestaltung der Maßnahmen**: bei der Entscheidung, welche Maßnahmen zur Sicherstellung der KI-Kompetenz ergriffen werden, ist miteinzubeziehen, über welche technischen Kenntnisse, Erfahrung und Ausbildung die zu schulenden Personen verfügen. Die Schulungsmaßnahmen können daher je nach Organisation und Kontext variieren.

Research Institute: Die neuen Verpflichtungen gemäß AI Act ab Februar 2025: Wie sich Organisationen jetzt optimal vorbereiten können

Gibt es spezifische Mindestanforderungen an die Schulungsinhalte (z. B. technische oder regulatorische Aspekte)?

Spezifische Mindestanforderungen sieht der AI Act nicht vor, dennoch können Inhalte von zu vermittelnden Kompetenzen etwa sein:

- **Kenntnis der jeweiligen Unternehmensstrategie, sowie angrenzender Richtlinien, sowie der Ansprechpersonen**
- **Basiswissen digitaler Kompetenz, etwa nach DigiComp 2.3 AT (https://www.digitalekompetenzen.gv.at/kompetenzen/Kompetenzmodell.html)**
  - **Digitalisierung in der Arbeitswelt**
  - **Informationen suchen und kritisch hinterfragen**
  - **Digitale Technologien für den Arbeitsalltag nutzen**
- **Verständnis von KI und deren Anwendungsbereich**
  - **Funktionsweise und Beispiele für KI-Einsatz**
  - **Innovationsmöglichkeiten durch KI, einhergehende Arbeitserleichterungen**
  - **Besonderheiten der Arbeitsweise von KI, insbesondere etwa hinsichtlich Bias, Halluzinationen, Bedeutung von Trainingsdaten**
- **Hinweise zur Nutzung der konkret im Unternehmen eingesetzten KI-Systeme**
- **Nutzerschulungen, Möglichkeiten zum Erfahrungsaustausch, unternehmensinterne Best Practices**
- **Etwa: Prompting-Workshops beim Einsatz von Text-KI**
- **Für verbreitete KI-Systeme bietet hier die Wirtschaftskammer eine gute Übersicht: https://www.wko.at/digitalisierung/ki-loesungen-fuer-die-praxis**
- **Ethische Aspekte**
- **Rechtliche Aspekte**
  - **Kenntnis der relevanten Aspekte aus Datenschutz, Patientenrechte, Arbeitsrecht, Urheberrecht, etc.**

Best Practices sind etwa:

- **Regelmäßig wiederkehrende Erhebung von eingesetzten KI-Systemen. Regelmäßige Re-Evaluierung von potentiellen KI-Use-Cases basierend auf dem Stand der Technik.**
- **Interdisziplinare Beobachtung durch Vertreter:innen aus Technik, Recht, Compliance, IT-Security, Human Ressources und Betriebsrat – abhängig von der Unternehmensgröße.**

Praxisorientiertes Lernen: Regelmäßiges Testen und Evaluieren neuer Systeme in Hinblick auf unternehmensinterne Use-Cases und Feedback einholen "KI-Kompetenz"

Literacy - Durchsetzbarkeit

Technology

Art 86 KI-VO (AI Act) verankert das Recht auf Erläuterung der Entscheidungsfindung im Einzelfall. Dh, dass von einer KI-gestützten Entscheidung betroffenen Personen ein Auskunftsanspruch in Bezug auf die Hintergründe und Mechanismen der Entscheidungsfindung zukommt.

Dabei ist es jedoch von immenser Bedeutung, dass Unternehmen und Vereine, vor allem aber auch öffentliche Einrichtungen, im Sinne der Compliance mit den Prinzipien der Transparenz und Rechenschaftspflicht diesem Betroffenenrecht in geeigneter Weise nachkommen, zumal der Auskunftsanspruch für Betroffene oftmals die Voraussetzung darstellt, um auf Basis dieser Informationen bestimmte Ansprüche wie etwa Schadenersatz geltend machen bzw den Rechtsweg beschreiten zu können.

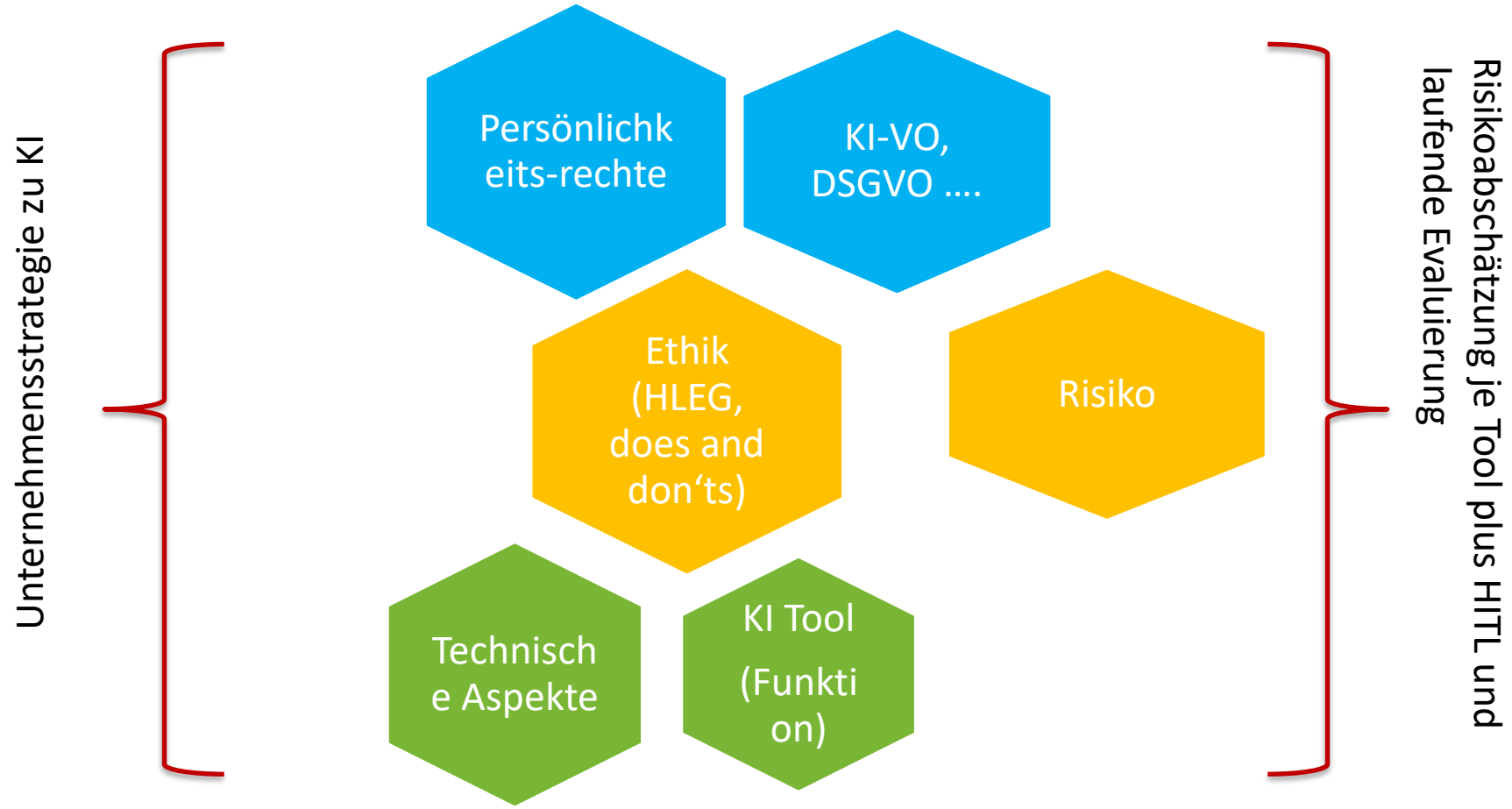
Bei der Analyse des Auskunftsanspruchs nach Art 86 KI-VO ist auch die Zusammenschau mit rechtlich relevanten datenschutzrechtlichen Regelungen, insbesondere Art 22 DSGVO von Bedeutung.

Neben den rechtlichen Verpflichtungen, hat sich das Erfordernis der Einhaltung von ethischen Prinzipien zu einem zentralen Eckpfeiler der österr KI-Strategie entwickelt.

Research Institute: Die neuen Verpflichtungen gemäß AI Act ab Februar 2025: Wie sich Organisationen jetzt optimal vorbereiten können

FH JOANNEUM
University of Applied Sciences

Umsetzung im Unternehmen



Information und Schulung für Mangement, MA, Partner, Kunden ..?

AI Act: Verpflichtungen von Betreibern

Der Umfang der Verpflichtungen nimmt entsprechend der Risikoklassifizierung des KI-Systems ab

	Hochrisiko KI-System	**KI-System** begrenzt. Risiko	**KI-System** minimal. Risiko
KI-Kompetenz	Art. 4	Art. 4	Art. 4
Transparenz gegenüber nachgelagerten Akteuren	Art. 26 (11)	Art. 50 (3), (4)	
Verwendung des KI-Systems laut Betriebsanleitung	Art. 26 (1), (3), (4)		
Menschliche Aufsicht	Art. 26 (2)		
Überwachung des KI-Systems	Art. 26 (5)		
Meldung von schwerwiegenden Vorfällen	Art. 26 (5), 73		
Aufbewahrung von erzeugten Protokollen	Art. 26 (6)		
Sofern relevant, Datenschutz-Folgenabschätzung	Art. 26 (9)		
Zusammenarbeit mit zuständigen nationalen Behörden	Art. 26 (12)		
Recht auf Erläuterung der Entscheidungsfindung im Einzelfall	Art. 86 (1)		
Informationspflichten gegenüber der Arbeitnehmer:innen-Vertretung *sofern Arbeitgeber:in Hochrisiko-KI-Systeme am Arbeitsplatz einsetzt*	Art. 26 (7)		
Registrierungspflicht *sofern EU-Organe, EU-Einrichtungen und sonstige EU-Stellen*	Art. 26 (8), 49		
Genehmigungspflicht einer Justiz- oder Verwaltungsbehörde *sofern Einsatz zur nachträglichen biometrischen Fernidentifizierung*	Art. 26 (10)		
Erstellung einer Grundrechte-Folgenabschätzung *sofern u. a. öffentl. oder private Einrichtungen öffentliche Dienste erbringen*	Art. 27		

AI Act: Verpflichtungen von Anbietern

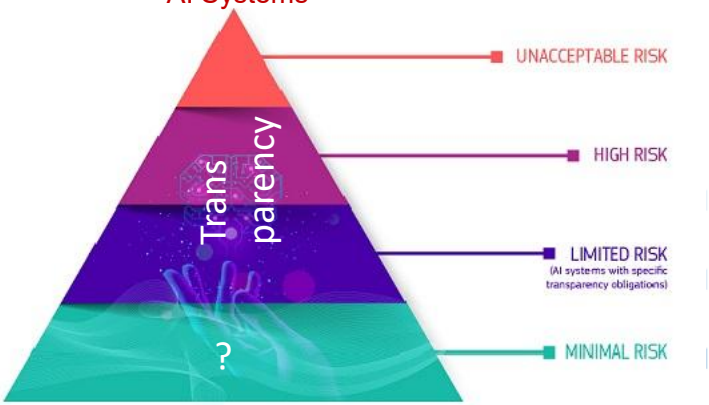
Der Umfang der Verpflichtungen nimmt entsprechend der Risikoklassifizierung des KI-Systems/KI-Modells ab

	Hochrisiko KI-System	**GPAI-Modell** system. Risiko	**GPAI-Modell**	**KI-System** begrenzt. Risiko	**KI-System** minimal. Risiko
KI-Kompetenz	Art. 4	Art. 4	Art. 4	Art. 4	Art. 4
Transparenz gegenüber nachgelagerten Akteuren	Art. 13	Art. 55 (1)	Art. 53 (1) b	Art. 50 (1), (2)	
Anforderungen an Daten	Art. 10	Art. 55 (1)	Art. 53 (1) c, d		
Technische Dokumentation	Art. 11	Art. 55 (1)	Art. 53 (1) a		
Zusammenarbeit mit Behörden	Art. 21	Art. 55 (1)	Art. 53 (3)		
Bennenung Bevöllmächtigter (sofern Drittstaat)	Art. 22	Art. 55 (1)	Art. 54		
Risikomanagement	Art. 9	Art. 55 (1) a, b			
Genauigkeit, Robustheit und Cybersicherheit	Art. 15	Art. 55 (1) d			
Registrierungs- bzw. Mitteilungspflichten	Art. 49	Art. 52 (1)			
Meldepflichten gegenüber Behörden	Art. 73	Art. 55 (1) c			
Aufzeichnung von Ereignissen	Art. 12				
Implementierung menschlicher Überwachungstools	Art. 14				
Kennzeichnungspflichten	Art. 16 b				
Sicherstellung der Barrierefreiheitsanforderungen	Art. 16 l				
Qualitätsmanagement	Art. 17				
Aufbewahrungspflichten	Art. 18, 19				
Korrekturmaßnahmen	Art. 20				
Konformitäts-Bewertungsverf., -Erklärung, -Kennzeichnung	Art. 43, 47, 48				

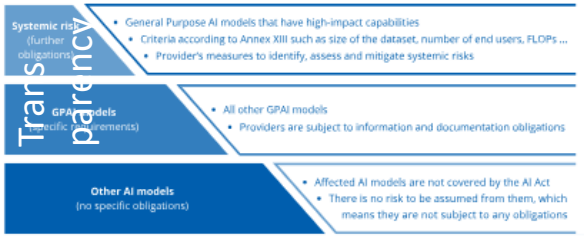
FH JOANNEUM
University of Applied Sciences

Transparency – providers, deployers (Art 13, Art 50, Art 53 AIA)

AI Systems



AI Models



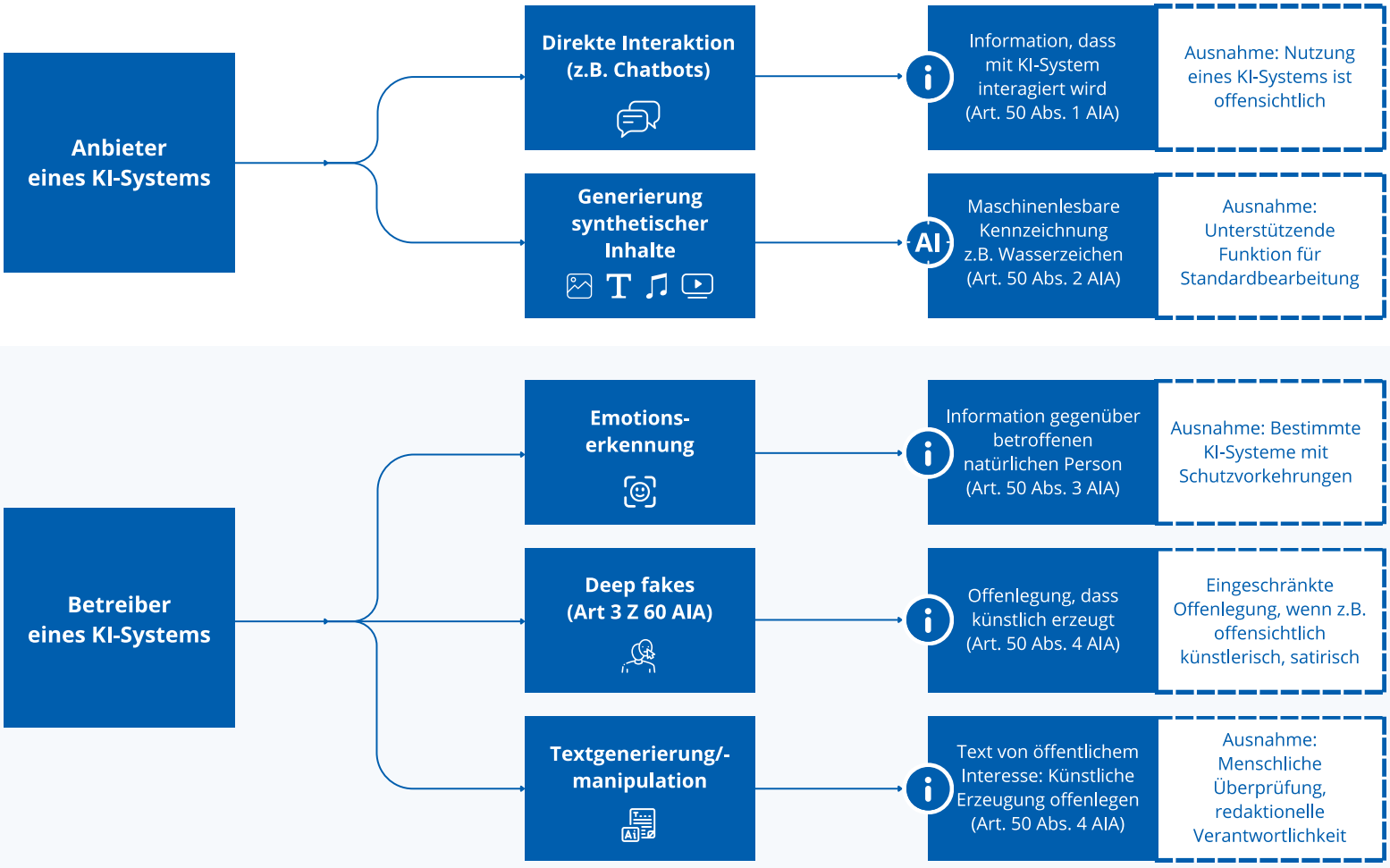
® RTR Austrian Regulatory Authority for Broadcasting and Telecommunications

Transparency: Affects primarily provider and deployer.
This means: Information, disclosure and labelling obligations, documentation, interpret andevaluate use and output

Branley-Bell / Proßnegg / Feiner / Libal, IRIS 2025, Vienna

AI Act: Transparenzpflichten

KI-Systeme mit begrenztem Risiko lösen Informations-, Offenlegungs- und Kennzeichnungspflichten aus



FH JOANNEUM
University of Applied Sciences

Transparency

Article 13

Transparency and provision of information to deployers

1. High-risk AI systems shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent t enable deployers to interpret a system's output and use it appropriately. An appropriate type and degree of transparency shall be ensure with a view to achieving compliance with the relevant obligations of the provider and deployer set out in Section 3.

2. High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise complete, correct and clear information that is relevant, accessible and comprehensible to deployers.

3. The instructions for use shall contain at least the following information:

Transparency: KI Entscheidung muss transparent sein, kontrolliert (kontrollierbar), und soll damit Vertrauen schaffen.

Branley-Bell / Proßnegg / Feiner / Libal, IRIS 2025, Vienna

FH JOANNEUM
University of Applied Sciences

The human in the / Transperancy

Abs 1 direkt interaction
Abs 2
Sog. Watermarking
Abs 3 emotion recognition, biometric categorisation
Abs 4 Deep Fake

CHAPTER IV

TRANSPARENCY OBLIGATIONS FOR PROVIDERS AND DEPLOYERS OF CERTAIN AI SYSTEMS

Article 50

Transparency obligations for providers and deployers of certain AI systems

1. Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system, unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate or prosecute criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, unless those systems are available for the public to report a criminal offence.

2. Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated. Providers shall ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art, as may be reflected in relevant technical standards. This obligation shall not apply to the extent the AI systems perform an assistive function for standard editing or do not substantially alter the input data provided by the deployer or the semantics thereof, or where authorised by law to detect, prevent, investigate or prosecute criminal offences.

3. Deployers of an emotion recognition system or a biometric categorisation system shall inform the natural persons exposed thereto of the operation of the system, and shall process the personal data in accordance with Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680, as applicable. This obligation shall not apply to AI systems used for biometric categorisation and emotion recognition, which are permitted by law to detect, prevent or investigate criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, and in accordance with Union law.

4. Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect,

Branley-Bell / Proßnegg / Feiner / Libal, IRIS 2025, Vienna

Legal

The human in the
Transperancy

Abs 1 direkt interaction
Abs 2
Sog. Watermarking
Abs 3 emotion recognition, biometric categorisation
Abs 4 Deep Fake

CHAPTER IV

TRANSPARENCY OBLIGATIONS FOR PROVIDERS AND DEPLOYERS OF CERTAIN AI SYSTEMS

Article 50

Transparency obligations for providers and deployers of certain AI systems

1. Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system, unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate or prosecute criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, unless those systems are available for the public to report a criminal offence.

2. Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated. Providers shall ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art, as may be reflected in relevant technical standards. This obligation shall not apply to the extent the AI systems perform an assistive function for standard editing or do not substantially alter the input data provided by the deployer or the semantics thereof, or where authorised by law to detect, prevent, investigate or prosecute criminal offences.

3. Deployers of an emotion recognition system or a biometric categorisation system shall inform the natural persons exposed thereto of the operation of the system, and shall process the personal data in accordance with Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680, as applicable. This obligation shall not apply to AI systems used for biometric categorisation and emotion recognition, which are permitted by law to detect, prevent or investigate criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, and in accordance with Union law.

4. Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect,

Branley-Bell / Proßnegg / Feiner / Libal, IRIS 2025, Vienna

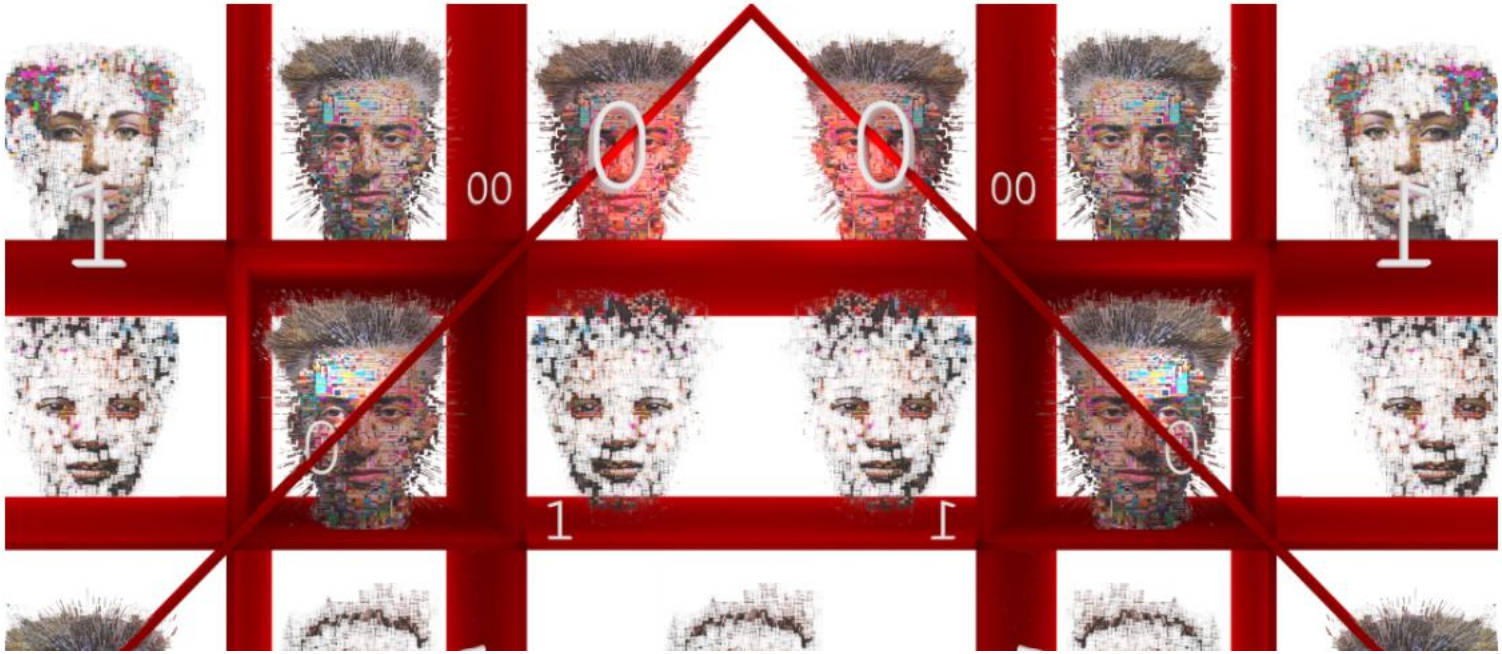
FH JOANNEUM
University of Applied Sciences

Transparency – Fehlentwicklungen …/ Verzweiflungsakte

NEWS

Denmark Leads EU Push to Copyright Faces in Fight Against Deepfakes

GIOVANA FLECK / OCT 7, 2025



AI Act: Verpflichtungen von Betreibern

Der Umfang der Verpflichtungen nimmt entsprechend der Risikoklassifizierung des KI-Systems ab

	Hochrisiko KI-System	**KI-System** begrenzt. Risiko	**KI-System** minimal. Risiko
KI-Kompetenz	Art. 4	Art. 4	Art. 4
Transparenz gegenüber nachgelagerten Akteuren	Art. 26 (11)	Art. 50 (3), (4)	
Verwendung des KI-Systems laut Betriebsanleitung	Art. 26 (1), (3), (4)		
Menschliche Aufsicht	Art. 26 (2)		
Überwachung des KI-Systems	Art. 26 (5)		
Meldung von schwerwiegenden Vorfällen	Art. 26 (5), 73		
Aufbewahrung von erzeugten Protokollen	Art. 26 (6)		
Sofern relevant, Datenschutz-Folgenabschätzung	Art. 26 (9)		
Zusammenarbeit mit zuständigen nationalen Behörden	Art. 26 (12)		
Recht auf Erläuterung der Entscheidungsfindung im Einzelfall	Art. 86 (1)		
Informationspflichten gegenüber der Arbeitnehmer:innen-Vertretung *sofern Arbeitgeber:in Hochrisiko-KI-Systeme am Arbeitsplatz einsetzt*	Art. 26 (7)		
Registrierungspflicht *sofern EU-Organe, EU-Einrichtungen und sonstige EU-Stellen*	Art. 26 (8), 49		
Genehmigungspflicht einer Justiz- oder Verwaltungsbehörde *sofern Einsatz zur nachträglichen biometrischen Fernidentifizierung*	Art. 26 (10)		
Erstellung einer Grundrechte-Folgenabschätzung *sofern u. a. öffentl. oder private Einrichtungen öffentliche Dienste erbringen*	Art. 27		

AI Act: Verpflichtungen von Anbietern

Der Umfang der Verpflichtungen nimmt entsprechend der Risikoklassifizierung des KI-Systems/KI-Modells ab

	Hochrisiko KI-System	**GPAI-Modell** system. Risiko	**GPAI-Modell**	**KI-System** begrenzt. Risiko	**KI-System** minimal. Risiko
KI-Kompetenz	Art. 4	Art. 4	Art. 4	Art. 4	Art. 4
Transparenz gegenüber nachgelagerten Akteuren	Art. 13	Art. 55 (1)	Art. 53 (1) b	Art. 50 (1), (2)	
Anforderungen an Daten	Art. 10	Art. 55 (1)	Art. 53 (1) c, d		
Technische Dokumentation	Art. 11	Art. 55 (1)	Art. 53 (1) a		
Zusammenarbeit mit Behörden	Art. 21	Art. 55 (1)	Art. 53 (3)		
Bennenung Bevöllmächtigter (sofern Drittstaat)	Art. 22	Art. 55 (1)	Art. 54		
Risikomanagement	Art. 9	Art. 55 (1) a, b			
Genauigkeit, Robustheit und Cybersicherheit	Art. 15	Art. 55 (1) d			
Registrierungs- bzw. Mitteilungspflichten	Art. 49	Art. 52 (1)			
Meldepflichten gegenüber Behörden	Art. 73	Art. 55 (1) c			
Aufzeichnung von Ereignissen	Art. 12				
Implementierung menschlicher Überwachungstools	Art. 14				
Kennzeichnungspflichten	Art. 16 b				
Sicherstellung der Barrierefreiheitsanforderungen	Art. 16 l				
Qualitätsmanagement	Art. 17				
Aufbewahrungspflichten	Art. 18, 19				
Korrekturmaßnahmen	Art. 20				
Konformitäts-Bewertungsverf., -Erklärung, -Kennzeichnung	Art. 43, 47, 48				

The human in the AI Act
Technological part – Literacy, Human oversight, the HITL

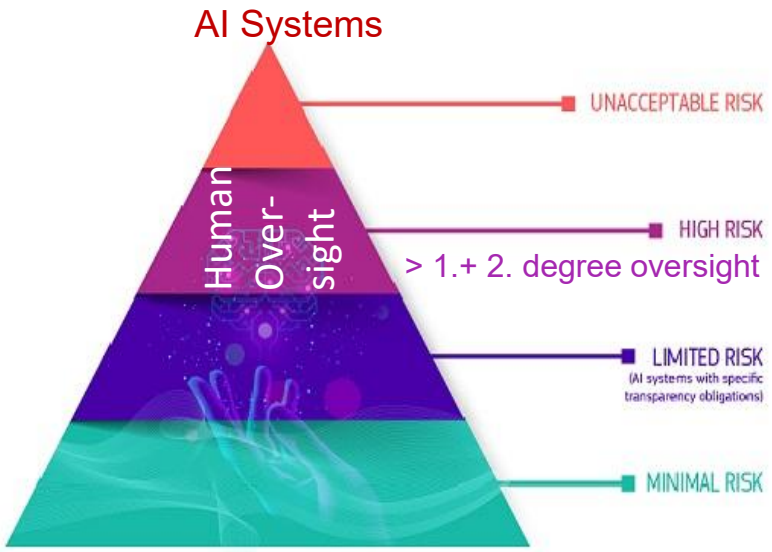
AI models often don't follow a clear, human-interpretable logic. Instead of a straightforward algorithm, they may operate through a highly parameterised model, where outputs are based on probabilities and statistical patterns. This raises issues around reliability, as outputs are often valid only across large data sets, leaving edge cases unaddressed or untested for accuracy. Even minor changes to input or training data, can lead to significant output variation. If applications update their models during use, it can create a system with inherent unpredictability, introducing states that are not easily reproducible. For users and developers alike, such systems incorporate logic of very high (unprecedented) complexity, often beyond human interpretation.

Reproducibility is also difficult, even impossible for some AI systems.

FH JOANNEUM
University of Applied Sciences

Legal

Human oversight



- ✓ Human Oversight Art 14 AIA
- ✓ Interface tool for effective oversight
- ✓ Prevent or minimize risks
- ✓ According to risk, autonomy, context of use
- ✓ Enables natural person to carry out an effective oversight

Human oversight: provider and deployer only for high risk AI-systems.

Branley-Bell / Proßnegg / Feiner / Libal, IRIS 2025, Vienna

FH JOANNEUM
University of Applied Sciences

The human oversight

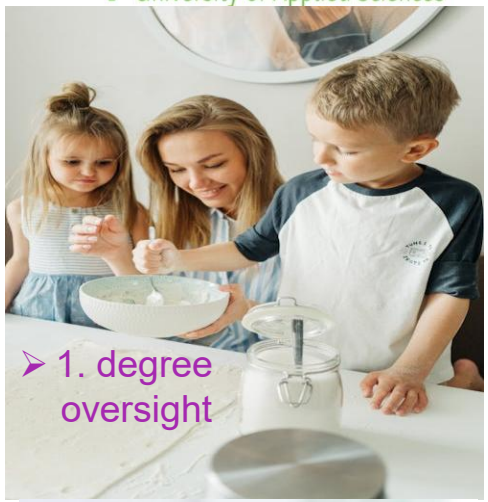
Legal

Oversight – When? Laux:
First-degree oversight involves the human directly in the decision making,

- level 1 as direct involvement of the human in the decision,
- level 2 as the use of AI to automate and save time with moderate engagement, and
- level 3 as the use of AI to (almost) replace human involvement in the decision-making process

Second-degree assigns the human to the role of reviewer, with the power of accepting or rejecting the AI decision/outcome.

- users may employ AI as a substitute for human effort
- in others they may use AI as a supplemental tool



> 2. degree oversight

Legal

The human in the AI Act Human Oversight

Article 14

Human oversight

1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use.

2. Human oversight shall aim to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular where such risks persist despite the application of other requirements set out in this Section.

3. The oversight measures shall be commensurate with the risks, level of autonomy and context of use of the high-risk AI system, and shall be ensured through either one or both of the following types of measures:

(a) measures identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;

(b) measures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the deployer.

4. For the purpose of implementing paragraphs 1, 2 and 3, the high-risk AI system shall be provided to the deployer in such a way that natural persons to whom human oversight is assigned are enabled, as appropriate and proportionate:

(a) to properly understand the relevant capacities and limitations of the high-risk AI system and be able to duly monitor its operation, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance;

(b) to remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;

(c) to correctly interpret the high-risk AI system's output, taking into account, for example, the interpretation tools and methods available;

(d) to decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output of the high-risk AI system;

(e) to intervene in the operation of the high-risk AI system or interrupt the system through a 'stop' button or a similar procedure that allows the system to come to a halt in a safe state.

5. For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 of this Article shall be such as to ensure that, in addition, no action or decision is taken by the deployer on the basis of the identification resulting from the system unless that identification has been separately verified and confirmed by at least two natural persons with the necessary competence, training and authority.

The requirement for a separate verification by at least two natural persons shall not apply to high-risk AI systems used for the purposes of law enforcement, migration, border control or asylum, where Union or national law considers the application of this requirement to be disproportionate.

FH JOANNEUM
University of Applied Sciences

The human in the AI Act: Human Oversight

Legal

Article 26

Obligations of deployers of high-risk AI systems

1. Deployers of high-risk AI systems shall take appropriate technical and organisational measures to ensure they use such systems in accordance with the instructions for use accompanying the systems, pursuant to paragraphs 3 and 6.

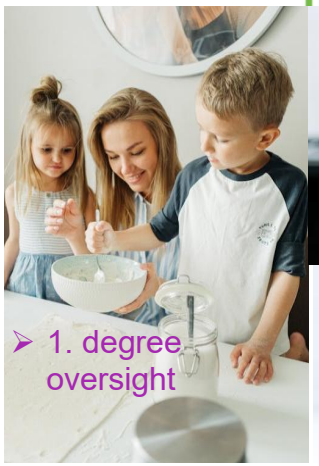
2. Deployers shall assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support.

3. The obligations set out in paragraphs 1 and 2, are without prejudice to other deployer obligations under Union or national law and to the deployer's freedom to organise its own resources and activities for the purpose of implementing the human oversight measures indicated by the provider.

4. Without prejudice to paragraphs 1 and 2, to the extent the deployer exercises control over the input data, that deployer shall ensure that input data is relevant and sufficiently representative in view of the intended purpose of the high-risk AI system.

5. Deployers shall monitor the operation of the high-risk AI system on the basis of the instructions for use and, where relevant, inform providers in accordance with Article 72. Where deployers have reason to consider that the use of the high-risk AI system in accordance with the instructions may result in that AI system presenting a risk within the meaning of Article 79(1), they shall, without undue delay

The human in the AI Act
Ethical considerations to the oversight



Oversight can be difficult and impractical

"Black box" AI systems hinder transparency and trust
It can be difficult even for those with an AI background to interpret how the system made its decision
Even more difficult to evaluate if the model was trained on data that included any errors or biases

Humans also show biases in how they interpret AI (e.g., overtrusting or undertrusting AI decisions). So human interpretation is not neutral either!

(For examples refer to Branley-Bell et al (2020), *User Trust and Understanding of Explainable AI: Exploring Algorithm Visualisations and User Biases. Doi: 10.1007/978-3-030-49065-2_27).*

FH JOANNEUM
University of Applied Sciences

Ethical part

The human in the AI Act
Ethical part

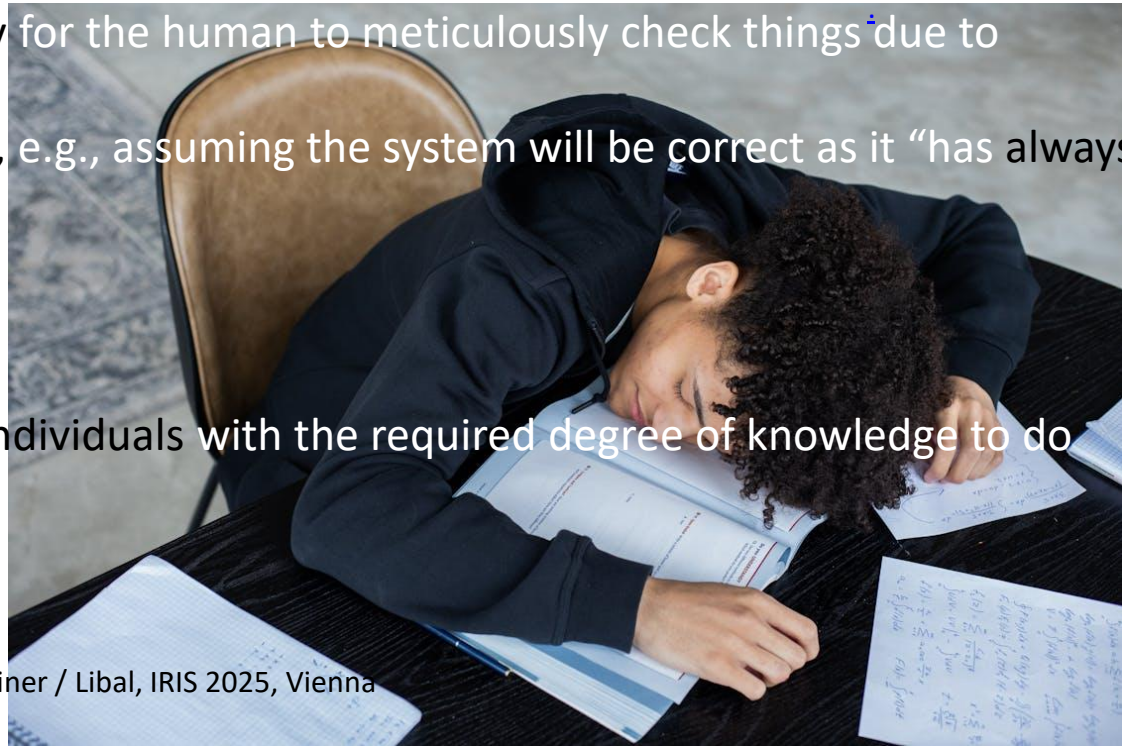
Oversight roles can be monotonous

- Can lead to reduced tendency for the human to meticulously check things due to boredom
- Can also lead to complacency, e.g., assuming the system will be correct as it "has always been correct previously"

Unrewarding for the human:

- Are we going to attract individuals with the required degree of knowledge to do tasks?

Pexels.com: Monstera Production

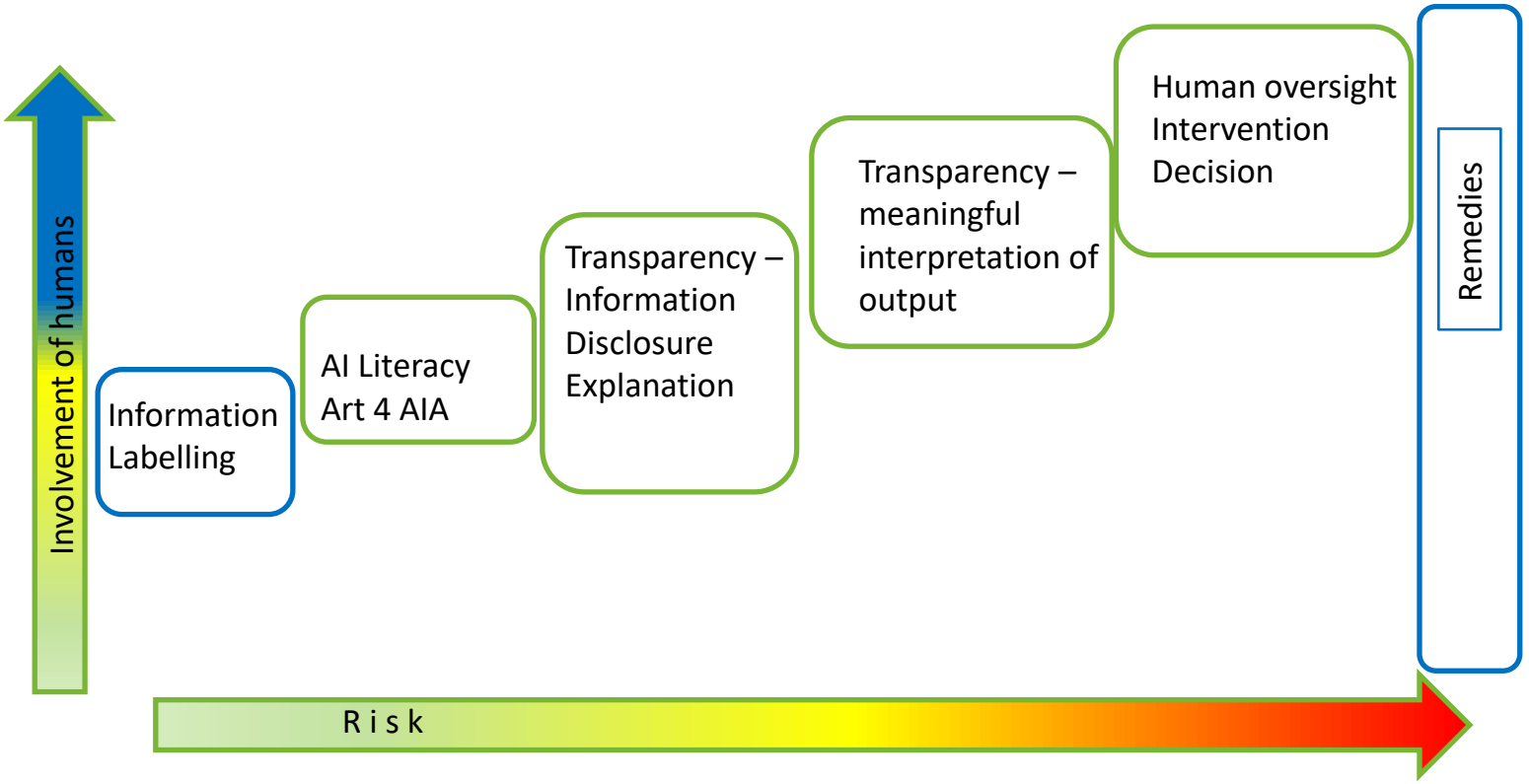


Branley-Bell / Proßnegg / Feiner / Libal, IRIS 2025, Vienna

FH JOANNEUM
University of Applied Sciences

Legal

Ladder for human involvement according to the level of risk and knowledge



Branley-Bell / Proßnegg / Feiner / Libal, IRIS 2025, Vienna

AI Act - Strafen

Bei folgenden Verstößen können folgende maximalen Geldbußen verhängt werden (siehe Art. 99, 101 AIA):

- **Bis zu 35 Mio. EUR oder 7 Prozent des gesamten weltweiten Vorjahresumsatzes (je nachdem, welcher Wert höher ist) bei Missachtung der verbotenen Praktiken;**
- **Bis zu 15 Mio. EUR oder 3 Prozent des gesamten weltweiten Vorjahresumsatzes (je nachdem, welcher Wert höher ist) bei Verstößen gegen Verpflichtungen, welche Konformitätsbewertungsstellen und die jeweiligen Akteure einzuhalten haben;**
  - **Anbieter von Hochrisiko-KI-Systemen, KI-Systemen mit begrenztem Risiko, GPAI-Modelle;**
  - **Betreiber von Hochrisiko-KI-Systemen und KI-Systemen mit begrenztem Risiko;**
  - **Bevollmächtigter;**
  - **Einführer;**
  - **Händler;**
- **Bis zu 7,5 Mio. EUR oder 1,5 Prozent des gesamten weltweiten Vorjahresumsatzes (je nachdem, welcher Wert höher ist) bei Bereitstellung falscher, unvollständiger oder irreführender Angaben an Konformitätsbewertungsstellen oder zuständige nationale Behörden auf deren Auskunftsersuchen.**

Da die **Organe, Einrichtungen und sonstigen Stellen der EU** mit gutem Beispiel vorangehen sollten, werden auch sie den Vorschriften und möglichen Sanktionen unterworfen. Bei folgenden Verstößen können folgende maximalen Geldbußen verhängt werden (siehe Art. 100 AIA):

- **Bis zu 1,5 Mio. EUR bei Missachtung der verbotenen Praktiken;**
- **Bis zu 750 000 EUR bei Nichtkonformität des KI-Systems mit in dieser Verordnung festgelegten Anforderungen oder Pflichten.**

AI Act

Wer Geldbußen verhängen darf, hängt damit zusammen, wer die Aufsicht übertragen bekommen hat.

- Dem Grunde nach sind es nationale Behörden, die Geldbußen verhängen dürfen (siehe Art. 99 AIA).
- Im Falle von Anbietern von GPAI-Modellen darf die Kommission Geldbußen verhängen (siehe Art. 101 AIA),
- im Falle von Verstößen gegen den AIA durch Organe, Einrichtungen und sonstige Stellen der EU ist hierzu der Europäische Datenschutzbeauftragter befugt (siehe Art. 100 AIA).

Bei der Frage, **ob und/oder in welchem Umfang eine Geldbuße verhängt werden soll**, sollen unter anderem folgende Aspekte berücksichtigt werden (siehe Art. 99 Abs. 7 AIA):

•Art, Schwere und Dauer des Verstoßes und seiner Folgen, unter Berücksichtigung des Zwecks des KI-Systems sowie gegebenenfalls der Zahl der betroffenen Personen und des Ausmaßes des von ihnen erlittenen Schadens;

•ob demselben Akteur bereits von anderen Marktüberwachungsbehörden für denselben Verstoß Geldbußen auferlegt wurden oder von anderen Behörden für Verstöße gegen das Unionsrecht oder das nationale Recht Geldbußen auferlegt wurden, wenn diese Verstöße auf dieselbe Handlung oder Unterlassung zurückzuführen sind, die einen einschlägigen Verstoß gegen diese Verordnung darstellt;

•Größe, Jahresumsatz und Marktanteil des Akteurs, der den Verstoß begangen hat;

•Vorsätzlichkeit oder Fahrlässigkeit des Verstoßes;

•Im Falle von Organen, Einrichtungen und sonstigen Stellen der EU sind besondere Gründe zu erwägen (siehe näher Art. 100 Abs. 1 AIA).

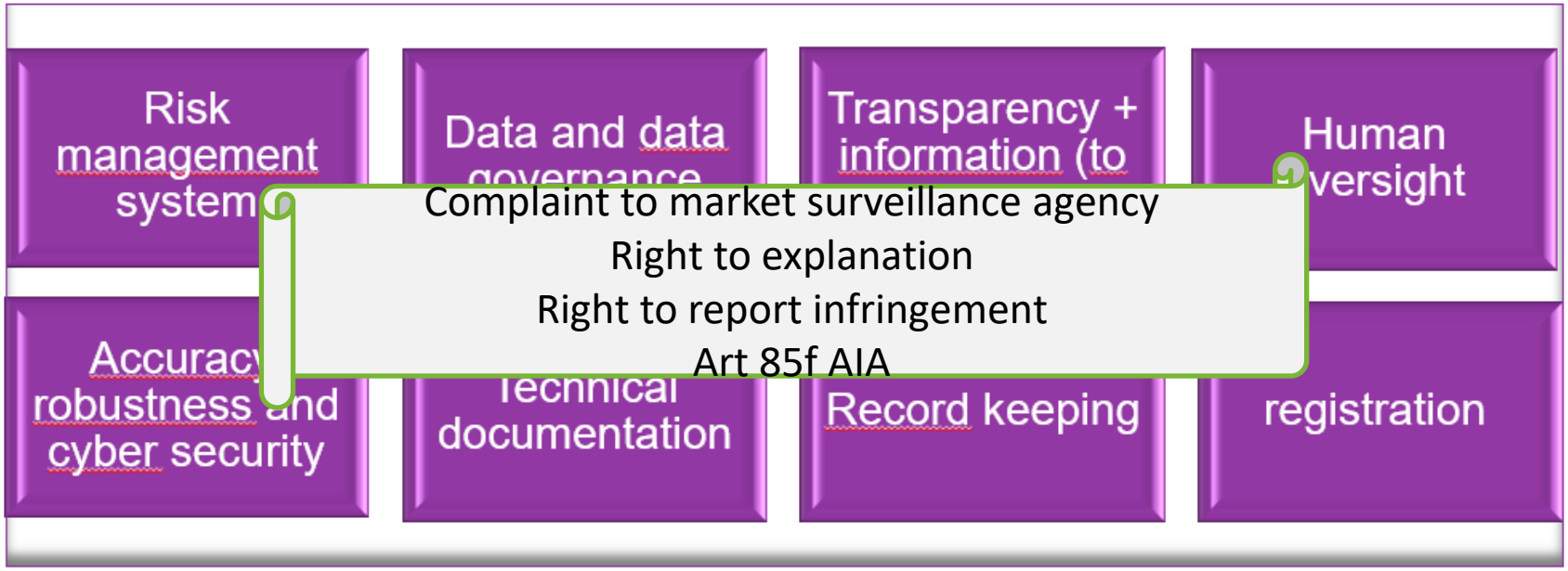
FH JOANNEUM
University of Applied Sciences

Exkurs: DSGVO

TOMs

FH JOANNEUM
University of Applied Sciences

Requirements for high risk AI systems and Remedies for humans



Kein umfassender Rechtsrahmen, sondern Produktsicherungsrecht für KI, ergänzend für bisherige Verbote, Transparenz und Individualrechtsschutz Vorgaben.
Relies mostly on state related agencies – registration, monitoring, …
Art 64ff AI Act: AI office, AI Board, Advisory Forum, Scientific Panel, Notifying Authority, Market Surveillance Authority; AI database and incidents database.

Rechtsmittel Remedies

European Union

EUR-Lex
Access to European Union law

Document 32024R1689

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance)

Article 85

Right to lodge a complaint with a market surveillance authority

Without prejudice to other administrative or judicial remedies, any natural or legal person having grounds to consider that there has been an infringement of the provisions of this Regulation may submit complaints to the relevant market surveillance authority.

In accordance with Regulation (EU) 2019/1020, such complaints shall be taken into account for the purpose of conducting market surveillance activities, and shall be handled in line with the dedicated procedures established therefor by the market surveillance authorities.

Article 86

Right to explanation of individual decision-making

1. Any affected person subject to a decision which is taken by the deployer on the basis of the output from a high-risk AI system listed in Annex III, with the exception of systems listed under point 2 thereof, and which produces legal effects or similarly significantly affects that person in a way that they consider to have an adverse impact on their health, safety or fundamental rights shall have the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken.

2. Paragraph 1 shall not apply to the use of AI systems for which exceptions from, or restrictions to, the obligation under that paragraph follow from Union or national law in compliance with Union law.

3. This Article shall apply only to the extent that the right referred to in paragraph 1 is not otherwise provided for under Union law.

Rechtsmittel

European Union

EUR-Lex

Access to European Union law

Legal

Document 32024R1689

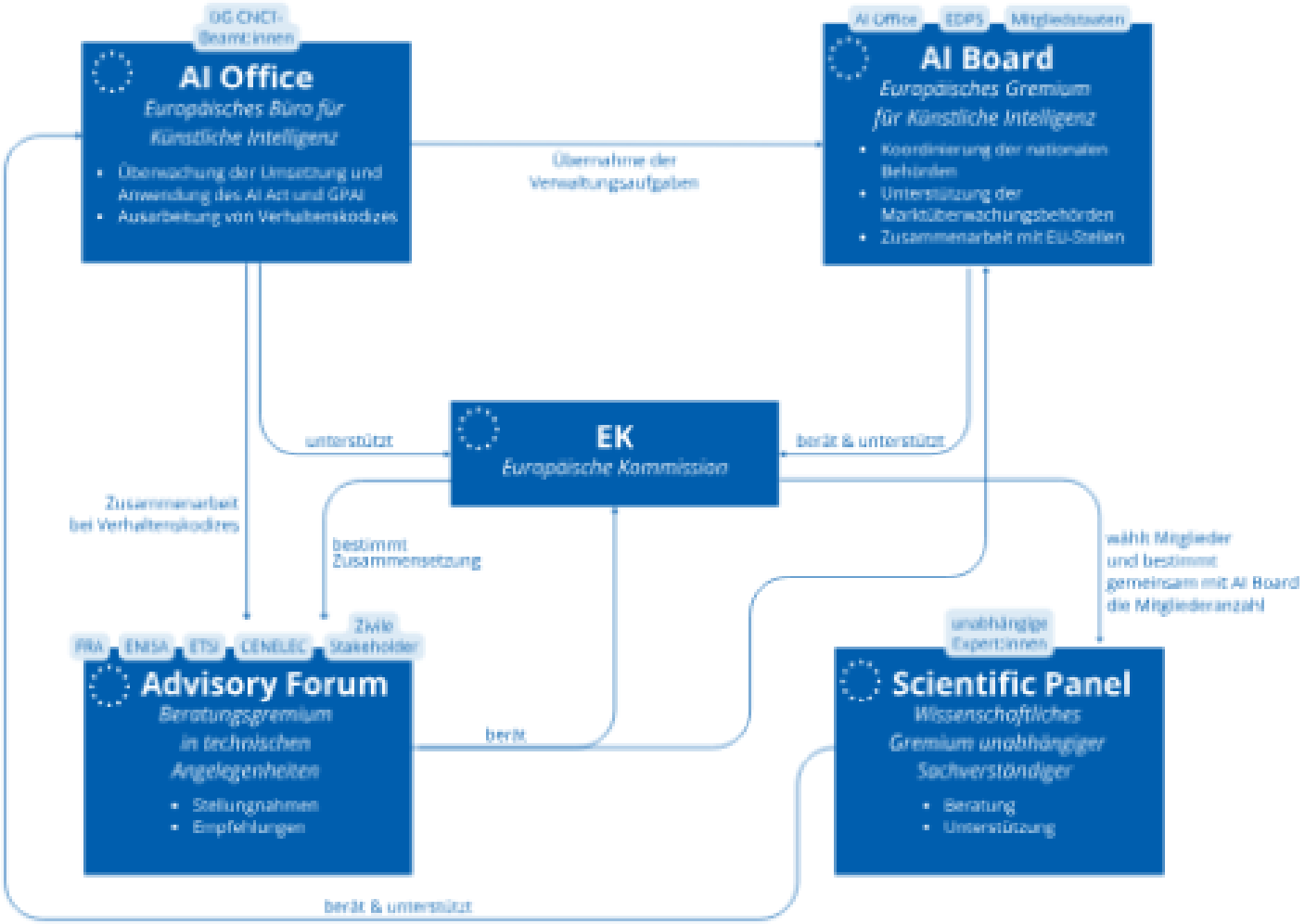
Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance)

(170) Union and national law already provide effective remedies to natural and legal persons whose rights and freedoms are adversely affected by the use of AI systems. Without prejudice to those remedies, any natural or legal person that has grounds to consider that there has been an infringement of this Regulation should be entitled to lodge a complaint to the relevant market surveillance authority.

(171) Affected persons should have the right to obtain an explanation where a deployer's decision is based mainly upon the output from certain high-risk AI systems that fall within the scope of this Regulation and where that decision produces legal effects or similarly significantly affects those persons in a way that they consider to have an adverse impact on their health, safety or fundamental rights. That explanation should be clear and meaningful and should provide a basis on which the affected persons are able to exercise their rights. The right to obtain an explanation should not apply to the use of AI systems for which exceptions or restrictions follow from Union or national law and should apply only to the extent this right is not already provided for under Union law.

AI Act: Behörden & Einrichtungen

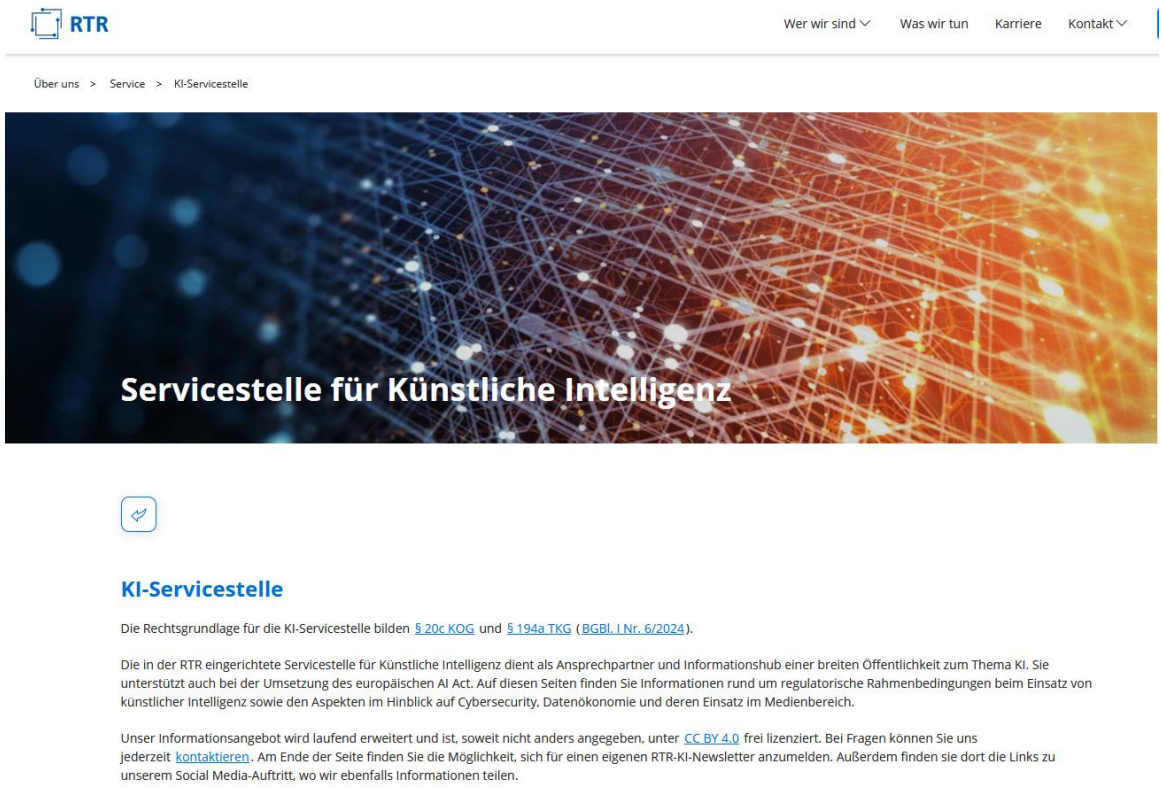
Welche Behörden und Einrichtungen gibt es? Für was sind sie zuständig?



Auf EU-Ebene sieht der AI Act vier Behördenstellen bzw. Einrichtungen vor. Die Infografik zeigt, wie sie sich zusammensetzen und welche Aufgaben sie übernehmen. © RTR (CC BY 4.0)

FH JOANNEUM
University of Applied Sciences

Behörde



European AI Office

PAGE CONTENTS

GenAI4EU

The Structure of the AI Office

Tasks of the AI Office

European Artificial Intelligence Board

Job opportunities and collaboration

The European AI Office is the centre of AI expertise across the EU. It plays a key role in implementing the AI Act - especially for general-purpose AI - fostering the development and use of trustworthy AI, and international cooperation.

The European AI Office supports the development and use of trustworthy AI, while protecting against AI risks. The AI Office was established within the European Commission as the centre of AI expertise and forms the foundation for a single European AI governance system.

Share

Quick Links

Artificial Intelligence

European approach to artificial intelligence

Coordinated Plan on Artificial Intelligence

AI Act

AI Pact

AI factories

Establishment of AI Office - press release

FH JOANNEUM
University of Applied Sciences

Behörde

Mitgliedstaat	Zuständige nationale Behörden(Art. 28 und Art. 70)	Grundrechte schützen(Artikel 77) Siehe auch Kommission konsolidiert Liste	Anmerkungen
Österreich	Unklar. Zur Unterstützung der Umsetzung des EU-KI-Gesetzes wurde bei der Rundfunk und Telekom Regulierungs-GmbH (RTR-GmbH) ein AI Service Desk eingerichtet. Die notifizierende Behörde und die Marktaufsichtsbehörde wurden nicht benannt.	Eine Liste von 19 Einrichtungen, die 8 verschiedene Bereiche abdecken, wurde von Digitales Österreich veröffentlicht.	Österreich hat drei Foren zu Unterstützung seiner KI-Politik eingerichtet: einen nationalen KI-Beirat, der si aus Experten aus Forschung und Wirtschaft zusammensetzt, das KI-Politikforum, das sich aus Mitgliedern verschiedener Ministerien zusammensetzt und das KI-Stakeholder-Forum, in dem verschieden Interessengruppen Beiträge leisten.

FH JOANNEUM
University of Applied Sciences

Dt Behörde

Bundesnetzagentur BNetzA

Bundesamt für die Sicherheit in der Informationstechnik (BSI)

Datenschutzbehörde

Neue Einrichtung

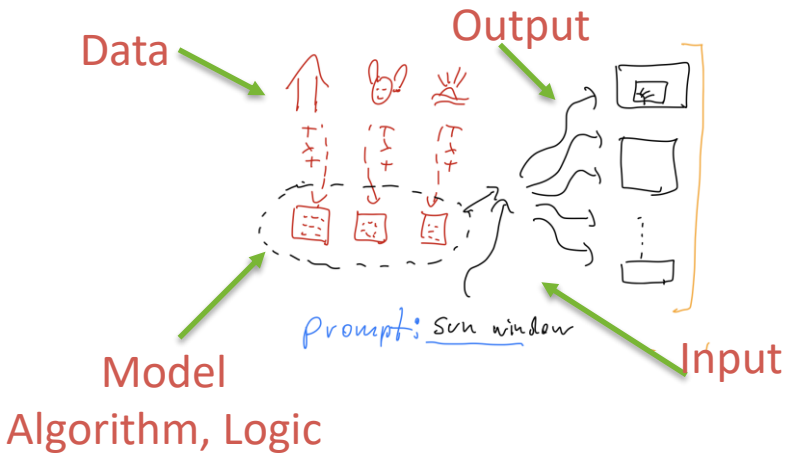
Attacks — Be Aware

Expect the evil:

Unintentionally biased/wrong

Intentionally biased/wrong

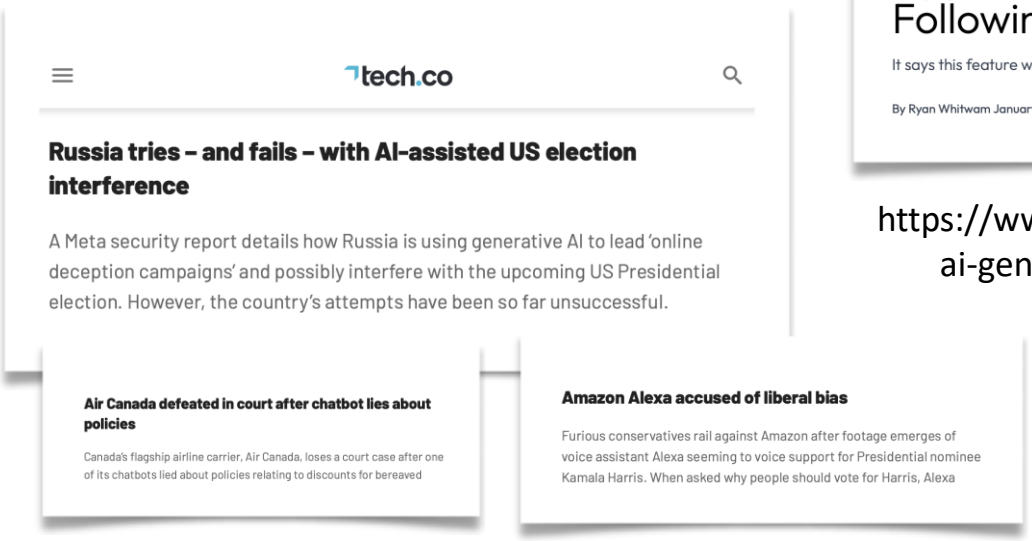
Cracking systems: steal data, modify data, modify models, intercept prompts,...



FH | JOANNEUM
University of Applied Sciences

Technology

Example/Problems



https://tech.co/news/list-ai-failures-mistakes-errors

https://www.extremetech.com/electronics/apple-pulls-ai-generated-news-summaries-following-errors

FH JOANNEUM
University of Applied Sciences

Technology

Was?

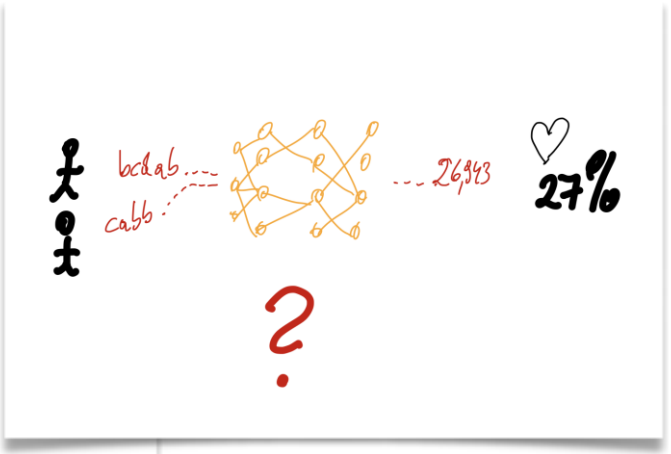
Literacy!

How does it work?

Consequences?

When to say No to AI?

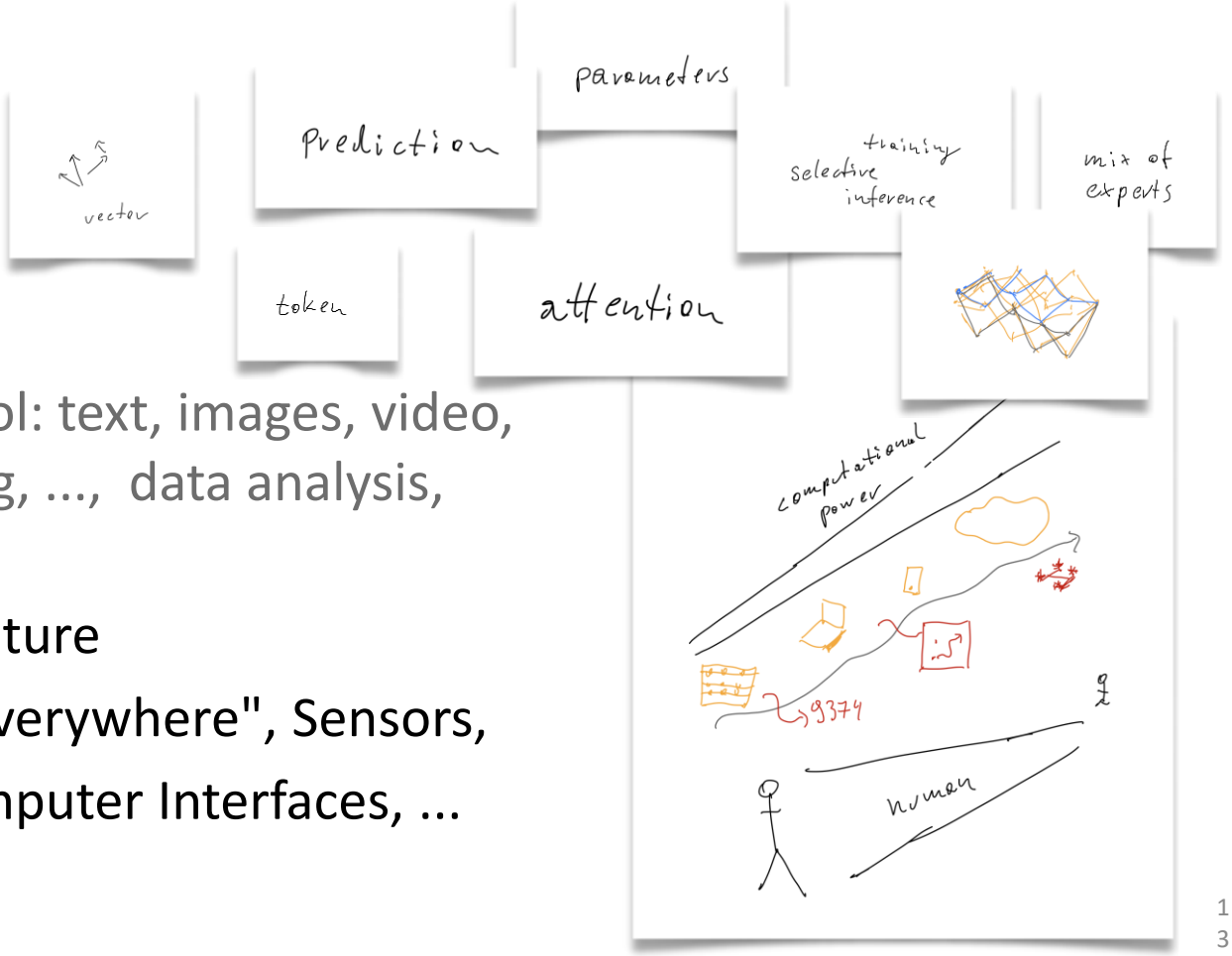
Whom to train?



FH JOANNEUM
University of Applied Sciences

Technology

Outlook



Currently

Useful "office" tool: text, images, video, ..., chats, ... coding, ..., data analysis, research ideas, ...

Near, future far future

Cheap, built-in "everywhere", Sensors,

Robots, Brain Computer Interfaces, ...

Beispiel: Firma Emotional AI (chin. UN) baut KI Emotionserkennung, AI Analytics (USA) vertreibt es, Siemens (Dl) verwendet es für AN

HLEG Recommendations

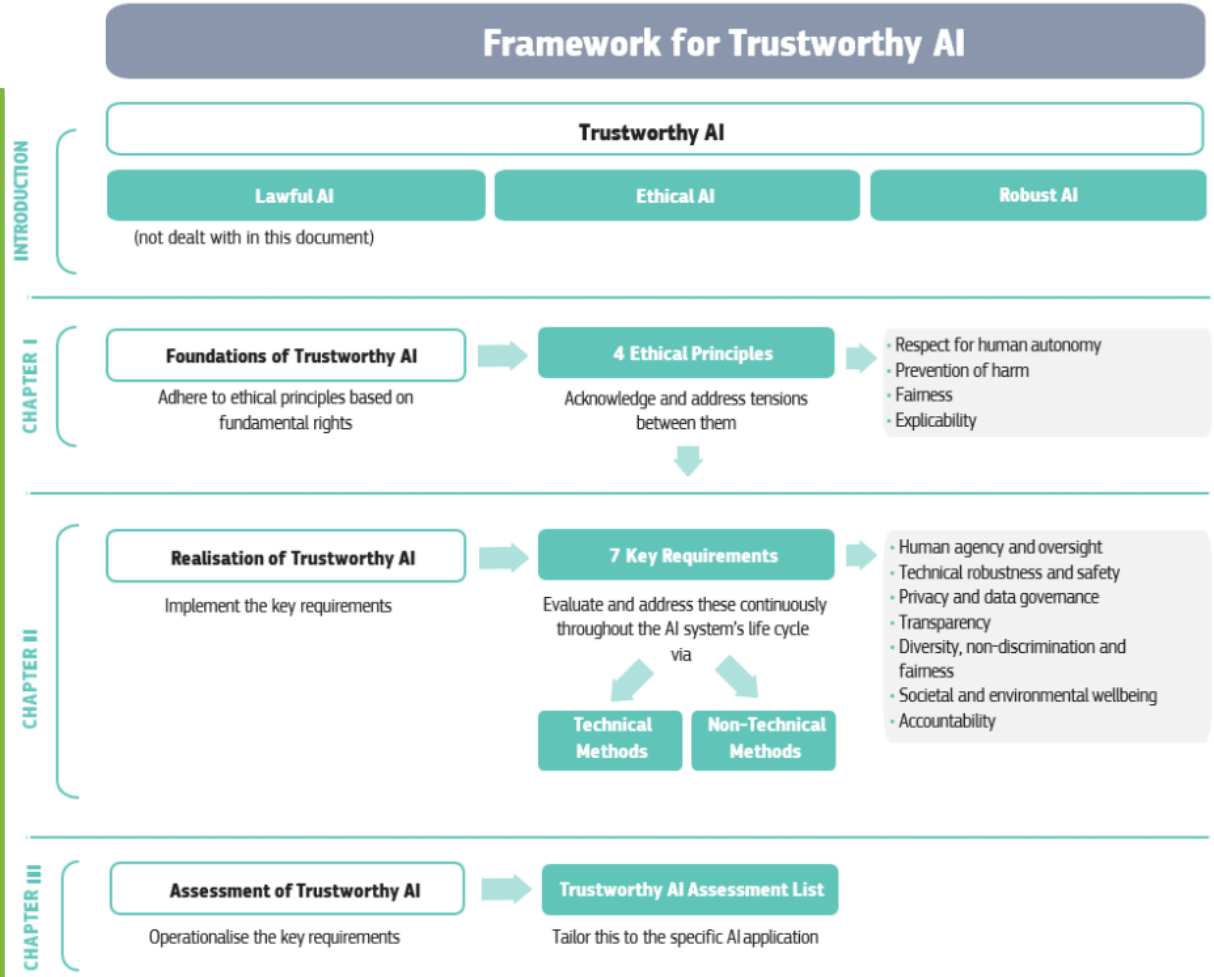


Figure 1: The Guidelines as a framework for Trustworthy AI

HLEG, Ethic Guidelines for Trustworthy AI, 2019

HLEG Recommendations

Gemäß den Leitlinien sollte eine vertrauenswürdige KI:
(1) rechtmäßig – Einhaltung aller geltenden Rechts- und Verwaltungsvorschriften
2) Ethik – Achtung ethischer Grundsätze und Werte
(3) robust – sowohl aus technischer Sicht als auch unter Berücksichtigung des sozialen Umfelds

FH JOANNEUM
University of Applied Sciences

HLEG Recommendations – 7 Kernanforderungen

Vorrang menschlichen Handelns und menschlicher Aufsicht

HLEG Recommendations – 7 Kernanforderungen

Technische Robustheit und Sicherheit

FH JOANNEUM
University of Applied Sciences

HLEG Recommendations – 7 Kernanforderungen

Privatsphäre und Datenqualitätsmanagement:

HLEG Recommendations – 7 Kernanforderungen

Transparenz:

FH JOANNEUM
University of Applied Sciences

HLEG Recommendations – 7 Kernanforderungen

Vielfalt, Nichtdiskriminierung und Fairness:

FH JOANNEUM
University of Applied Sciences

HLEG Recommendations – 7 Kernanforderungen

Gesellschaftliches und ökologisches Wohlergehen:

FH JOANNEUM
University of Applied Sciences

HLEG Recommendations – 7 Kernanforderungen

Rechenschaftspflicht:

FH | JOANNEUM
University of Applied Sciences

A Case Study of Tessa, an AI-Powered Chatbot

Dawn Branley-Bell, Associate Professor (Northumbria University, UK)
dawn.branley-bell@northumbria.ac.uk

Johannes Feiner, Senior Lecturer (FH JOANNEUM)
johannes.feiner@fh-joanneum.at

Sabine Proßnegg, Associate Professor (FH JOANNEUM)
sabine.prossnegg@fh-joanneum.at

Structure

- Part 1: Chatbots in mental health settings, an introduction to Tessa
- Part 2: What went wrong?
- Part 3: How can we prevent similar occurrences in future from a technological and user perspective, and from a legal perspective
- Conclusion

FH JOANNEUM
University of Applied Sciences

Chatbots

AI-based computer programmes designed to mimic human-to-human conversation by analysing the user's text-based input, and providing smart, related answers [Dahiya, 2017].

Chatbots give real time responses and the user can 'chat' to them like they would to another human.

The main benefits often cited are:

- Easy, *instant* access to services
- Reduced burden on staff
- Increased service capacity
- Reduced costs

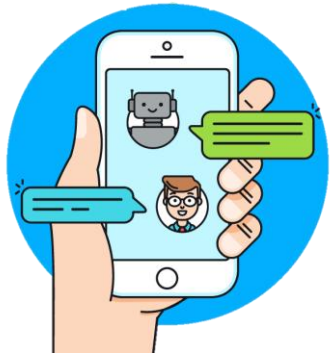
Common for online troubleshooting, generating quotes (e.g., insurance quotes), tracking deliveries etc. → and of course ChatGPT!

Dahiya, M. (2017). A tool of conversation: Chatbot. *International Journal of Computing Sciences and Engineering* 5, 158–161.

Chatbots in Mental Health

Mental health may seem like a risky or controversial area to introduce chatbots (e.g., concerns over reducing human services) but there are potential benefits:

- Increase healthcare capacity and reduced burden on overworked, fatigued healthcare staff. *Is a chatbot as an instant contact point better than nothing?*
- Improve access to healthcare:
  - Reduced wait times
  - Convenience - e.g., can access from home, flexible access times, no travel time or costs
  - Reduced barriers to access - e.g., constraints due to finances, location, caregiving responsibilities, comorbid health conditions
  - Reduced costs - controversial!
- Technology can be effective as a *complement* to, not a *replacement* for human care! [



FH JOANNEUM
University of Applied Sciences

Chatbots & Eating Disorders

Highest mortality rate of all mental health conditions	Delays to support access can be disastrous
Individuals rarely seek help (significant prevalence rates also a large underestimation!)	Need to improve possibility of reaching this hard-to-reach population
People often rely on the internet for help anyway **Digital tools can be some users' preference**	If professional support isn't available, they could turn to less helpful and/or inaccurate resources (misinformation, triggering pro-ana content etc)
Many EDs are comorbid with other mental health conditions	Can create additional barriers to access, including barriers to attending face-to-face support
EDs are linked to stigma and incorrect negative stereotypes and feelings of shame	Chatbots can potentially provide an encouraging 'first step' for individuals who feel embarrassed or stigmatised [Branley-Bell, 2023]

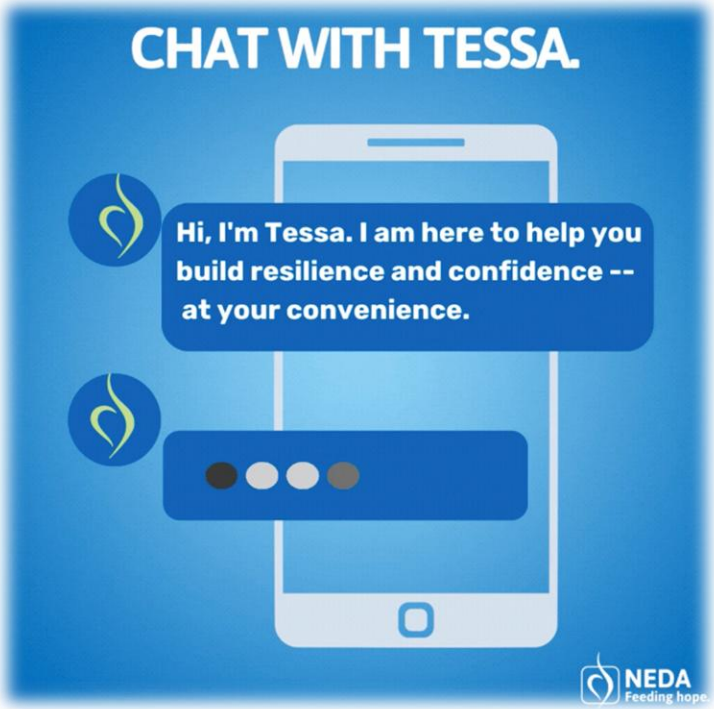
Branley-Bell et al. (2023). Chatbots for embarrassing and stigmatizing conditions: could chatbots encourage users to seek medical advice? *Frontiers in Communication* https://doi.org/10.3389/fcomm.2023.1275127

FH JOANNEUM
University of Applied Sciences

Who is Tessa?

A specialist wellness chatbot designed to support individuals vulnerable to eating disorders

- Designed by a researcher and then adopted by NEDA (National Eating Disorders Association)
- Officially launched in February 2022
- “Not designed to replace human interaction” but to act as a first point of contact to direct towards relevant support
- Also offered guidance on building resilience and confidence for individuals at very early stages of experiencing body image issues



Fitzsimmons-Craft, et al. (2022). “Effectiveness Of A Chatbot For Eating Disorders Prevention: A Randomized Clinical Trial”. *International Journal of Eating Disorders*. https://doi.org/10.1002/eat.23662

But…

The New York Times

A Wellness Chatbot Is Offline After Its 'Harmful' Focus on Weight Loss

The artificial intelligence tool, named Tessa, was presented by the National Eating Disorders Association as a way to discover coping skills. But activists say it instead veered into problematic weight-loss advice.

Psychiatrist.com CME JOURNALS NEWS

NEDA Suspends AI Chatbot for Giving Harmful Eating Disorder Advice

THE CUT STYLE | SELF | CULTURE

ARTIFICIAL INTELLIGENCE | JUNE 2, 2023

Turns Out Chatbots Aren't Great at Eating-Disorder Counseling

By Bindu Bansinath, a writer for the Cut who covers news, culture, and relationships.

An eating disorders chatbot offered dieting advice, raising fears about AI in health

UPDATED JUNE 9, 2023 · 6:59 AM ET

By Kate Wells

FROM Michigan Public

Tessa was terminated in May 2023

Many questions raised…

1. How did this happen – what went wrong?
2. What can we learn from it and how can we prevent similar occurrences in future –
   - From a technological perspective?
   - From a user perspective?
   - From a legal perspective?

What went wrong?

- Source code of Tessa not freely available
- Researcher behind the development of Tessa states that it was designed as a ‘rule-based system’ with a limited number of pre-determined responses

  → it should not have been able to go ‘off piste’ [Chan et al. 2022]
- Somehow Tessa switched to a generative language approach (similar to ChatGPT)

What went wrong?

AI systems…

- Are not always correct
- Can change over time
- Often lack transparency or explainability
- Do not promote reproducibility
- Can be vulnerable to malicious attacks
- May raise questions over privacy
- Can enforce the tendency towards monopolies

What can be done from a technological perspective to protect against occurrences like Tessa in future?

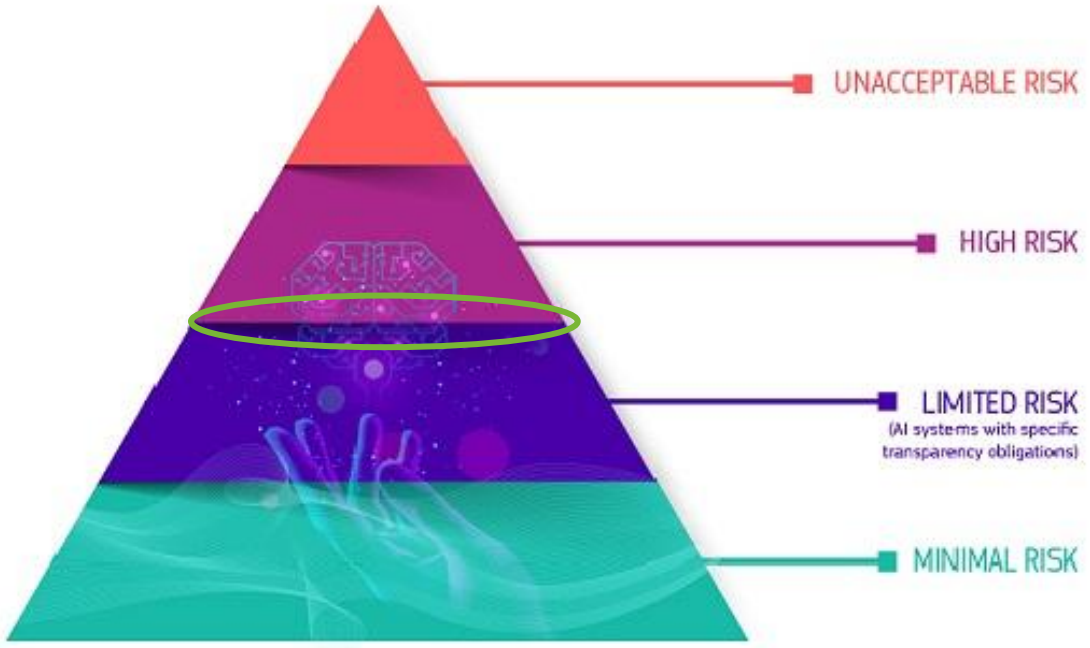
What can be done from a user perspective?

- Trust in technology is crucial if we are going to interact with tools and systems like chatbots for mental health
- BUT incidences like Tessa can greatly erode human trust and this can be difficult to rebuild [Lukyanenko et al., 2022]
- Interestingly, humans can also show a tendency to *overtrust* technology due to assuming automation is not vulnerable to bias or inaccuracies [Branley-Bell et al., 2020; Buçinca et al., 2021] – important to foster *accurate* trust (not under or over trust!)

→ Beneficial to build systems, and educate users, to enable critical reflection of how a system is operating – this requires transparency and explainability to enable human evaluate of system performance

FH | JOANNEUM
University of Applied Sciences

The risk-based approach and Tessa



Critical infrastructures (e.g. transport), that could put the life and health of citizens at risk
Tessa as part of health system?

Health application or

Just a wellness application?

Is (was) Tessa a mere chatbot?

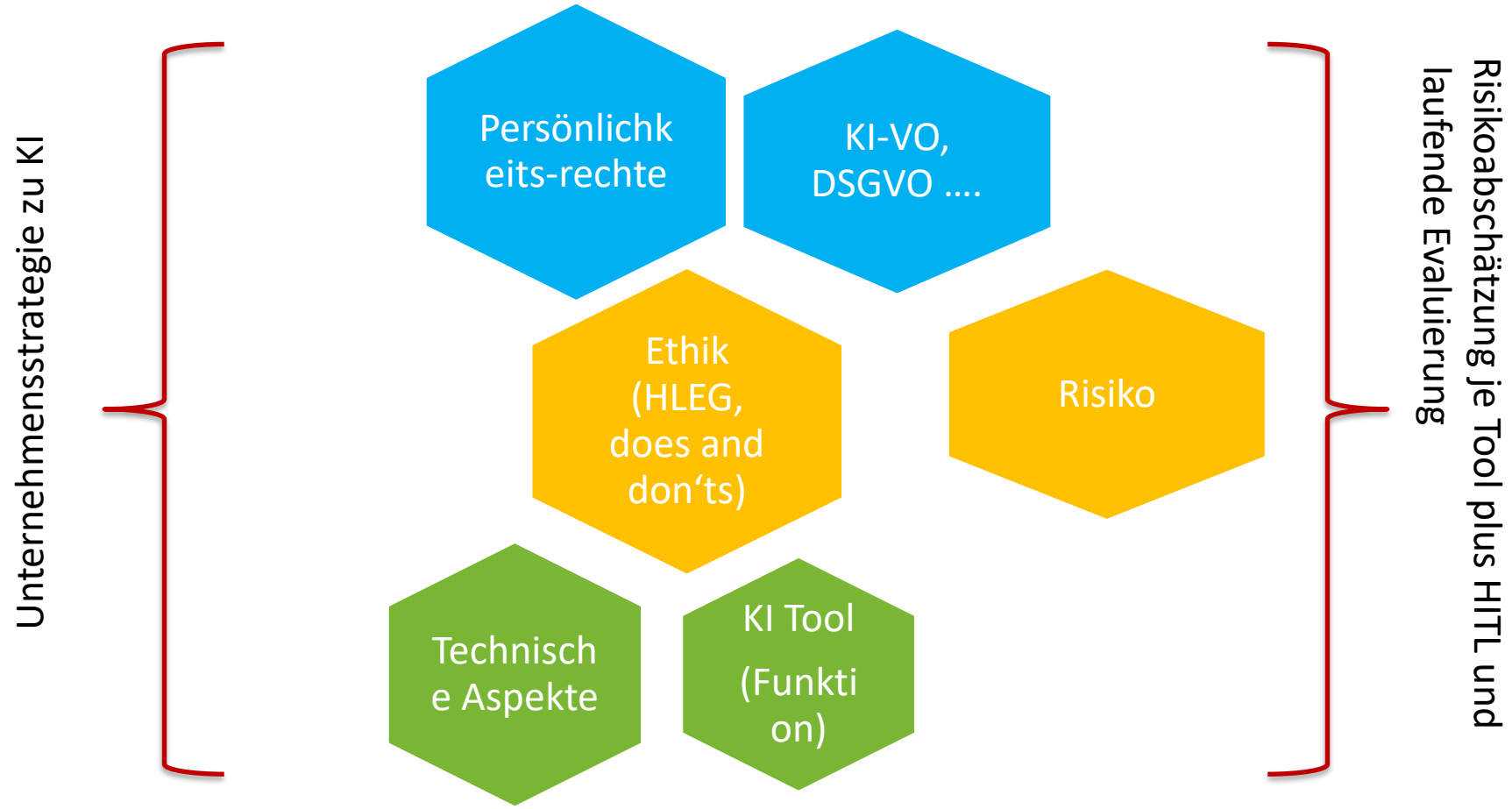
Risk assessments for AI applications

- GDPR
- OECD
- NIST
- High Level Expert Group, Trustworthy AI

-> risk assessment fatigue

FH JOANNEUM
University of Applied Sciences

Umsetzung im Unternehmen



Information und Schulung für Mangement, MA, Partner, Kunden ..?

FH JOANNEUM
University of Applied Sciences

How to be compliant?

- Strategien
- Umsetzungskonzepte
- Information
- Trainings
- Evaluierungen

- … Papier auch “leben”.

https://gesellschaft-datenschutz.de/die-10-ki-gebote/

KI –generierte Leistung ist keine eigene Leistung.

KI macht auch Fehler.

KI gibt und nimmt.

Schutz vor Data- und Textmining durch Bot-Sperren und § 44b dtUrhG

Prompts mit Personendaten (oder geschützten Daten wie Marken) nur mit Zustimmung.

An die Nutzungs- und Lizenzbedingungen der Unternehmen halten.

Achtung bei Schnittstellen (APIs) und Veröffentlichungen.

Achtung im HR Bereich.

Compiance
Regelmäßiger Update rechtlich, technologisch und faktisch.

https://gesellschaft-datenschutz.de/die-10-ki-gebote/

FH JOANNEUM
University of Applied Sciences

RADIUS

FFG
Promoting Innovation.

Research in Artificial Intelligence
for
Development, Innovation
and
Upgraded Security

Projekt RADIUS

Project Duration:

1.1.2025 bis 31.12.2029

Projekt Team:

appr. 30 Personen,
PL Helmut Lindner

Budget:

appr. € 1.7 Mio.

Workpackages:

AI-Systems

AI in Software Engineering

AI in Security

Use Cases

- Aviation /Luftfahrt
- Embedded Systems
- Banks and Insurances/
  Banken- und Versicherungswesen

https://www.pexels.com/ pixabay

APPLIED COMPUTER SCIENCES

FH JOANNEUM

DIHSÜD
Digital Innovation Hub

DIGITALISIERUNG IST EASY

Folgen Sie uns

für mehr Informationen zu Veranstaltungen, Digitalisierung und Innovation.

einfügen Partner-Kontaktdaten

einfügen Partner-Logo

THANK YOU!