TRENCHANT

Rechtsanwaltsgesellschaft mbH
RAin Marion Schultz

Webinar
IHK-München und Oberbayern
21.11.2024

AI-Act
Umsetzung beim privatrechtlichen Unternehmer-Betreiber

Julius Kirschbaum

Affiliationen:

FAU
Friedrich-Alexander-Universität
Fachbereich Wirtschafts- und Sozialwissenschaften | WiSo

Wissenschaftlicher Mitarbeiter im Bereich Wirtschaftsinformatik

INGENIEURE OHNE GRENZEN

Ehrenamtliche Tätigkeit im Bereich Weiterbildung und Kompetenzentwicklung

VDI Wissensforum

Leitung für den Zertifikatslehrgang Fachingenieur GenAI Sprachmodelle VDI

Themen

1. Einführung (warum)
   a) Grundrechtsausgleich
   b) Geschützte Rechtsgüter
   c) Grundsatz der Verhältnismäßigkeit: Risikobasierter Ansatz
2. Schutzmechanismus der KI-Verordnung (was)
   a) Übersicht
   b) KI-Kompetenz beim Betreiber („Compliance by Competence“)
   c) Dokumentation und Information
   d) Technische Anforderungen an Hochrisiko-KI-Systeme („Compliance by Design“)
3. Organisation (wie)
   a) Risikomanagement beim Betreiber
   b) Schulungen
4. Technische Grenzen der menschlichen Aufsicht und Kontrolle ⇒ **Julius Kirschbaum**
5. Inkrafttreten und Anwendbarkeit
6. Fragen und Diskussion

TRENCHANT

Einführung

Grundrechtsausgleich

- KI-Verordnung als „Grobkonzept“ eines angemessenen Interessenausgleichs auf Basis einer Grundrechtsabwägung
- Mittel zum angemessenen Ausgleich: Vertrauenswürdiger KI-Systeme
  - ➢ menschenzentrierte KI-Systeme
  - ➢ Menschliche Aufsicht
- Konkretisierung u.a. über “harmonisierte Normen“
- **Exkurs**: Zum Thema “Europa hängt sich ab“: KI-Konvention des Europarates v. 17.5.2024 unter Mitwirkung u.a. USA, Kanada, Japan, Australien

Bild: Adobe Stock

Geschützte prominente Grundrechte (Auszüge)

Was liegt in der Waagschale:

- Menschenwürde
  - Nicht als Objekt gesiebt, sortiert, bewertet, gruppiert, konditioniert oder manipuliert zu werden.
- Freiheit des Einzelnen
  - Freiheit von (in)direktem unrechtmäßigem Zwang, von Bedrohungen für die Selbstbestimmung, von ungerechtfertigter Überwachung, Täuschung und unfairer Manipulation
- Schutz der unternehmerischen Freiheit
- Achtung von Demokratie, Gerechtigkeit und Rechtsstaatlichkeit
- Nichtdiskriminierung
- Leben und körperliche Unversehrtheit
- Sichere Zukunft durch Innovationskraft / Wissenschaft und Forschung
- Privatsphäre und informationelle Selbstbestimmung

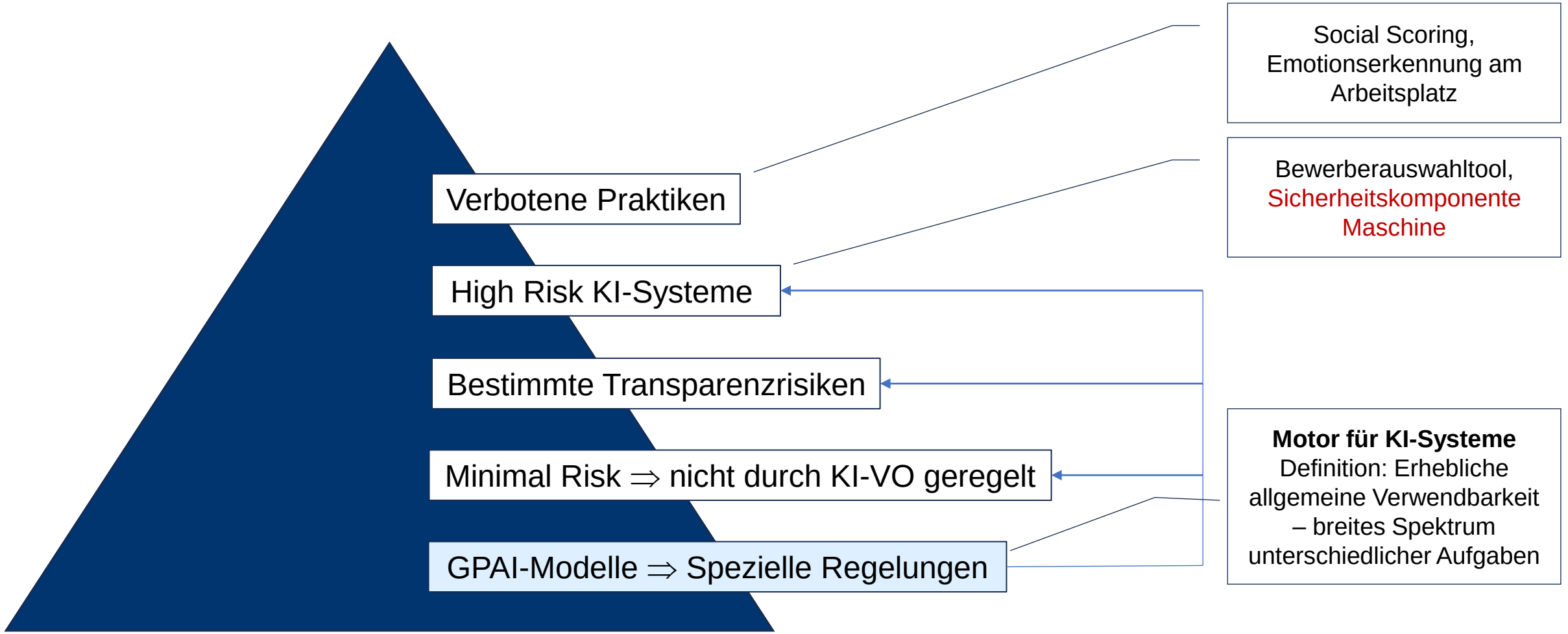
Effizienzsteigerung
Produktentwicklung

Kreditvergabe
Personalrekrutierung

Autonome Fahrzeuge
Medizinischer Einsatz

KI-Systeme – Risikoabhängige Regulierung

Wie schwer wiegen sie

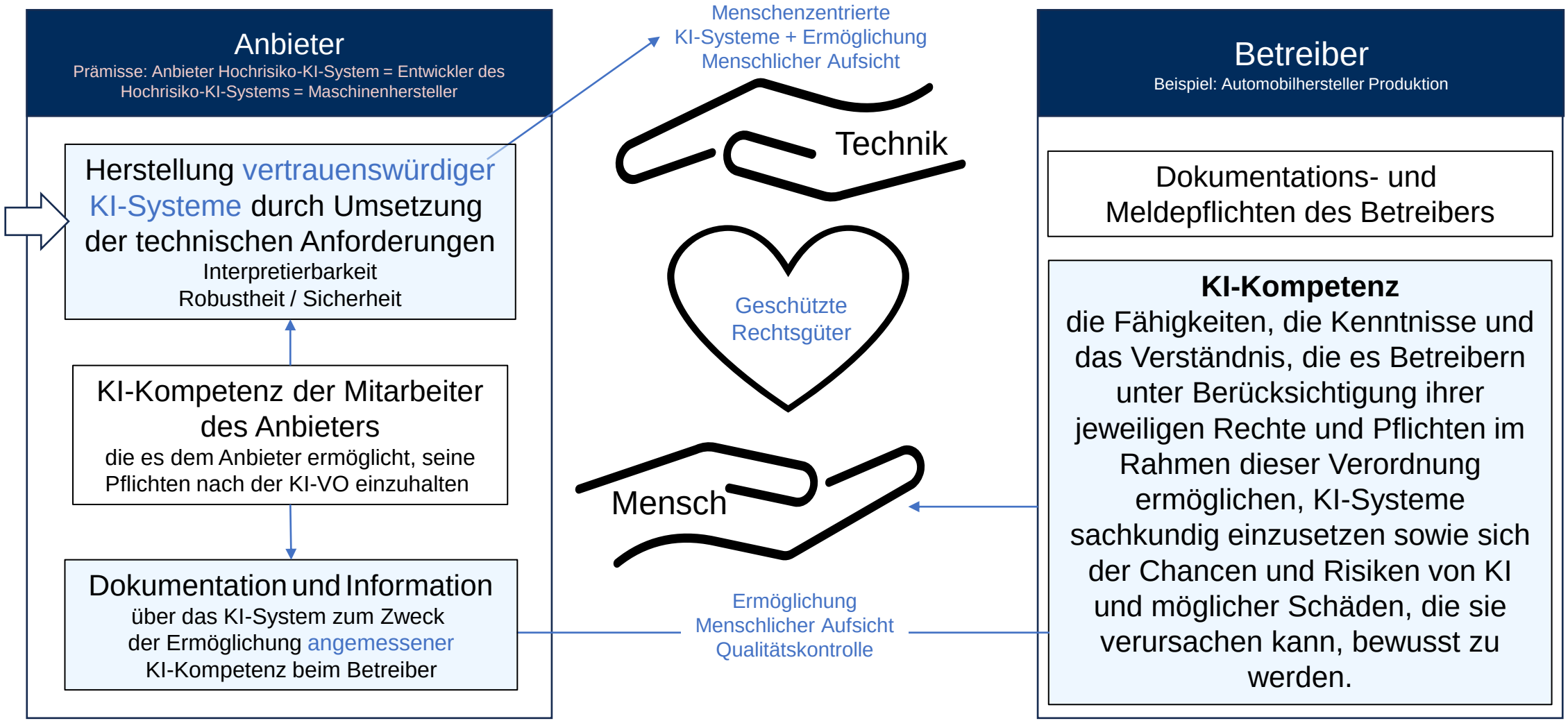


TRENCHANT

Schutzmechanismus der KI-Verordnung

Schutzmechanismus – Hochrisiko-KI-Systeme

Beispiel: Produktionsmaschine mit KI-Sicherheitsbauteil = Hochrisiko-KI-System



TRENCHANT

KI-Kompetenz beim Betreiber

Human oversight

KI-Kompetenz

Definition Art. 3 Nr. 56 KI-VO:
Für die Zwecke dieser Verordnung bezeichnet der Ausdruck „KI-Kompetenz“ die Fähigkeiten, die Kenntnisse und das Verständnis, die es Anbietern, Betreibern und Betroffenen unter Berücksichtigung ihrer jeweiligen Rechte und Pflichten im Rahmen dieser Verordnung ermöglichen, KI-Systeme sachkundig einzusetzen sowie sich der Chancen und Risiken von KI und möglicher Schäden, die sie verursachen kann, bewusst zu werden.

Zweck der KI-Kompetenz beim Betreiber:

- Sicherstellung der „menschlichen Aufsicht“
- Auswahl eines KI-Systems, das die technischen Anforderungen erfüllt

Benannte Aufsichtsperson ⟺ „KI-Beauftragter“

↓ KI-VO

↓ Ggf. aus allgemeinen Sorgfaltspflichten HGB/GmbHG/AktG

Konkretisierung der KI-Kompetenz beim Betreiber

Alle Mitarbeiter sollten folgende Kompetenzen erwerben:

- Grundlagenkompetenz zu KI-Systemen, ihren Möglichkeiten und Risiken.
- Rechtliche Grundlagenkompetenz KI-Verordnung und Produktsicherheitsrecht.

Die vom Betreiber benannte Person muss folgende Kompetenzen erwerben,
wenn Hochrisiko-KI-Systeme eingesetzt werden:

- Genaue Kenntnis des konkreten Verwendungskontext ⇒ Systembezogen
- Kenntnis der wahrscheinlich betroffenen Personen
- Erkennen erheblicher potenzieller Risiken
  - Kenntnis der geschützten Rechtsgüter („Schutzzweck der Norm“)
  - Kenntnis wichtiger Parameter zur Wahrung des „Human Oversight“.
    - Anforderungen der KI-Verordnung
    - Verstehen der Funktionsweise des KI-Systems auf Basis der Dokumentationen und Informationen des Anbieters

Dokumentation und Information

des Anbieters für den Betreiber zum Zweck

der Gewinnung von KI-Kompetenz

(„Schulungsunterlagen“)

Dokumentation GPAI-Modell

Technische Dokumentation des GPAI-Modells ohne systemische Risiken (Auszüge):

- Allgemeine Beschreibung des KI-Modells einschließlich:
  - Aufgaben, die das GPAI-Modell erfüllen soll
  - Art der KI-Systeme, in die es integriert werden kann
  - Anwendbaren Regelungen der akzeptablen Nutzung
- Informationen zum Entwicklungsprozess, einschließlich
  - Technische Mittel, die für die Integration des GPAI-Modells in KI-Systeme erforderlich sind (z.B. Betriebsanleitungen, Infrastruktur, Instrumente)
  - Modalität (z.B. Text oder Bild) und Format der Ein- und Ausgaben und deren maximale Größe
  - Informationen über die für das Trainieren, Testen und Validieren verwendeten Daten einschließlich der Art und Herkunft und der Aufbereitungsmethoden

! Vorsicht WebCrawling und UrhG

Technische Dokumentation Hochrisiko-KI-System

Technische Dokumentation des Hochrisiko-KI-Systems (Auszüge):

- Allgemeine Beschreibung des KI-Systems inkl. Zweckbestimmung
- Detaillierte Beschreibung der Bestandteile des KI-Systems und seines Entwicklungsprozesses, darunter
  - Methoden und Schritte zur Entwicklung
  - Allgemeine Logik des KI-Systems und der Algorithmen
  - Welche Bedeutung den verschiedenen Parametern zukommt
  - Trainingsmethoden und -techniken
  - Trainingsdaten
    - Herkunft, Umfang und Hauptmerkmale
    - Beschaffung und Auswahl, Kennzeichnungsverfahren und Datenbereinigungsmethoden
- Maßnahmen zur Ermöglichung der menschlichen Aufsicht und der Interpretation der Ergebnisse
- Detaillierte Informationen über Validierungs- und Testverfahren und die Genauigkeit

Folie 17

TRENCHANT

Technische Anforderungen an Hochrisiko-KI-Systeme

Interpretierbarkeit

Robustheit und Sicherheit

Technische Anforderungen

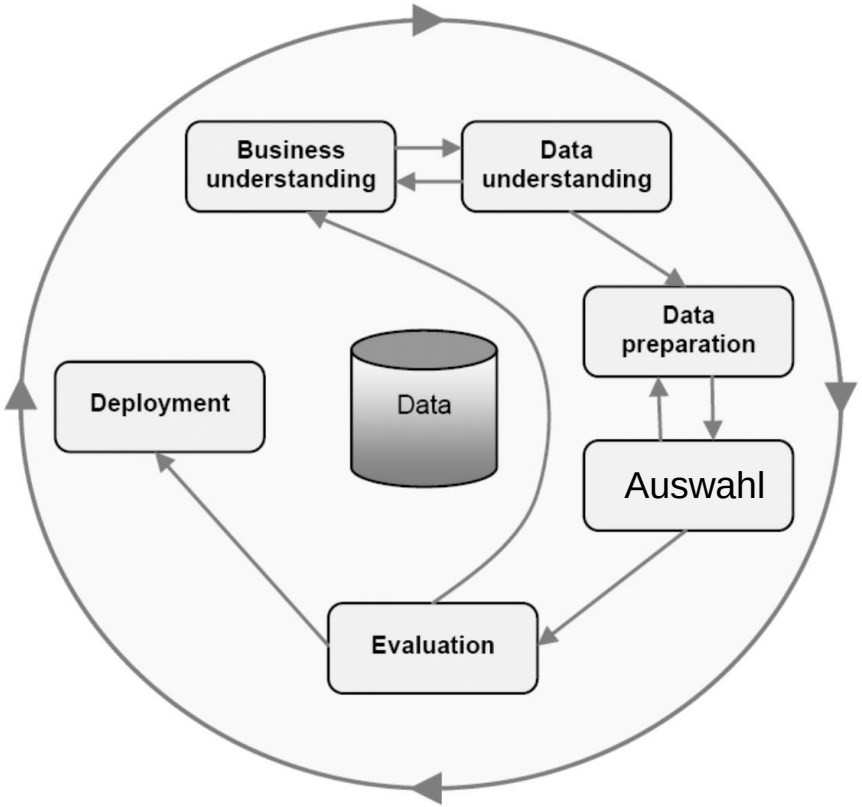
- **Interpretierbarkeit** durch Erklärbarkeit und Transparenz (⇒ Vertrauenswürdigkeit)
  - **Erklärbarkeit**: Der Grad, zu dem ein System Klarheit über die Gründe für die Ergebnisse verschaffen kann ⇒ **Technische Erklärbarkeit und Nachvollziehbarkeit** kann durch die Reproduzierbarkeit von Ergebnissen unterstützt werden. **Sachliche Erklärbarkeit** kann nur manchmal durch spezielle Methoden erlangt werden.
  - **Transparenz**: Der Grad, bis zu dem ein System Informationen über sein Innenleben, also seine innere Struktur und Trainingsdaten, offenbart ⇒ Öffnen der Blackbox. Entscheidungsbaum vs. Deep learning.
    - Transparenz by Design
    - Nachträgliche Herstellung von Transparenz durch bestimmte Methoden ⇒ Julius
- Den Menschen muss bewusst sein, dass sie mit einem KI-System kommunizieren oder interagieren - Kennzeichnungspflicht.
- Technische Robustheit und Sicherheit: KI-Systeme müssen widerstandsfähig sein gegen Versuche, die Verwendung oder Leistung des KI-Systems so zu verändern, dass dadurch die unrechtmäßige Verwendung durch Dritte ermöglicht und dadurch geschützten Rechtsgüter verletzt werden.

TRENCHANT

Organisation

(wie soll ich es machen)

Organisationspflicht Risikomanagement Betreiber



Do and document am Beispiel der Verbindung des KI-Systems mit einer eigenen Datenbasis (RAG)

≠ Risikomanagementsystem des **Anbieters** von Hochrisiko-KI-Systemen!

(1) Business Understanding: Verstehen des Anwendungsfalls und des Mehrwerts; Benennung einer Aufsichtsperson wenn Hochrisiko-KI-System

(2) Data Understanding: Auswahl der eigenen Datenbasis, die an das KI-Modell angeflanscht werden soll

(3) Data Preparation: Bereinigung und Aufbereitung der eigenen Datenbasis

(4) Auswahl des richtigen KI-Systems und Integration in eigene Datenbasis

(5) Evaluation: Test und der Bewertung der Ergebnisse.

(6) Deployment: Entscheidung zur Freigabe = Sorgfältige Delegation von Aufgaben.

(7) laufende Überwachung und Beachtung der „Gebrauchs-anweisung“.

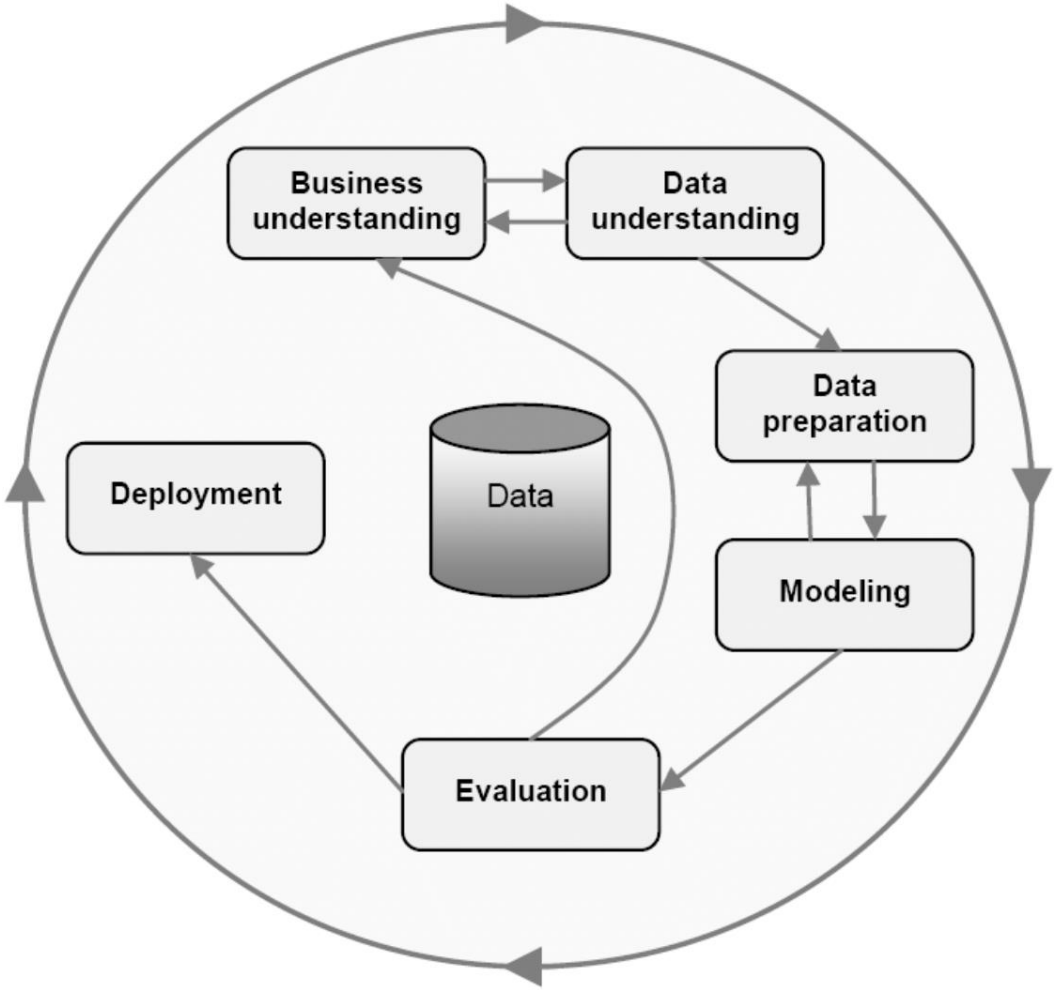
Organisationspflicht Schulungen

- Schulungsprogramme entwickeln und implementieren, die die jeweils notwendigen technischen und rechtlichen Grundlagen für den Umgang mit KI-Systemen vermitteln.
  - **Technische Schulung** zu KI-Systemen
  - **Rechtliche Schulung** zu KI-Verordnung (horizontale Gesetzgebung) und zu sektoralen Rechtsakten
- Trainings zur sicheren Nutzung von KI-Systemen bereitstellen, insbesondere wenn diese in sicherheitskritischen oder sensiblen Bereichen eingesetzt werden.
- Regelmäßige Auffrischungskurse für die Mitarbeitenden anbieten, um sicherzustellen, dass sie stets über die neuesten Entwicklungen und Best Practices im Umgang mit KI-Systemen informiert sind.
- Daneben: Besondere Schulung der benannten Person auf das konkrete KI-System

Technische Grenzen der Menschlichen Aufsicht und Kontrolle

Sicherstellung von Human Oversight über Systematik

Das CRISP-DM Modell



Wirth, R., and Hipp, J. 2000. "CRISP-DM: Towards a Standard Process Model for Data Mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29-39," *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining* (24959), pp. 29–39. (https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining).

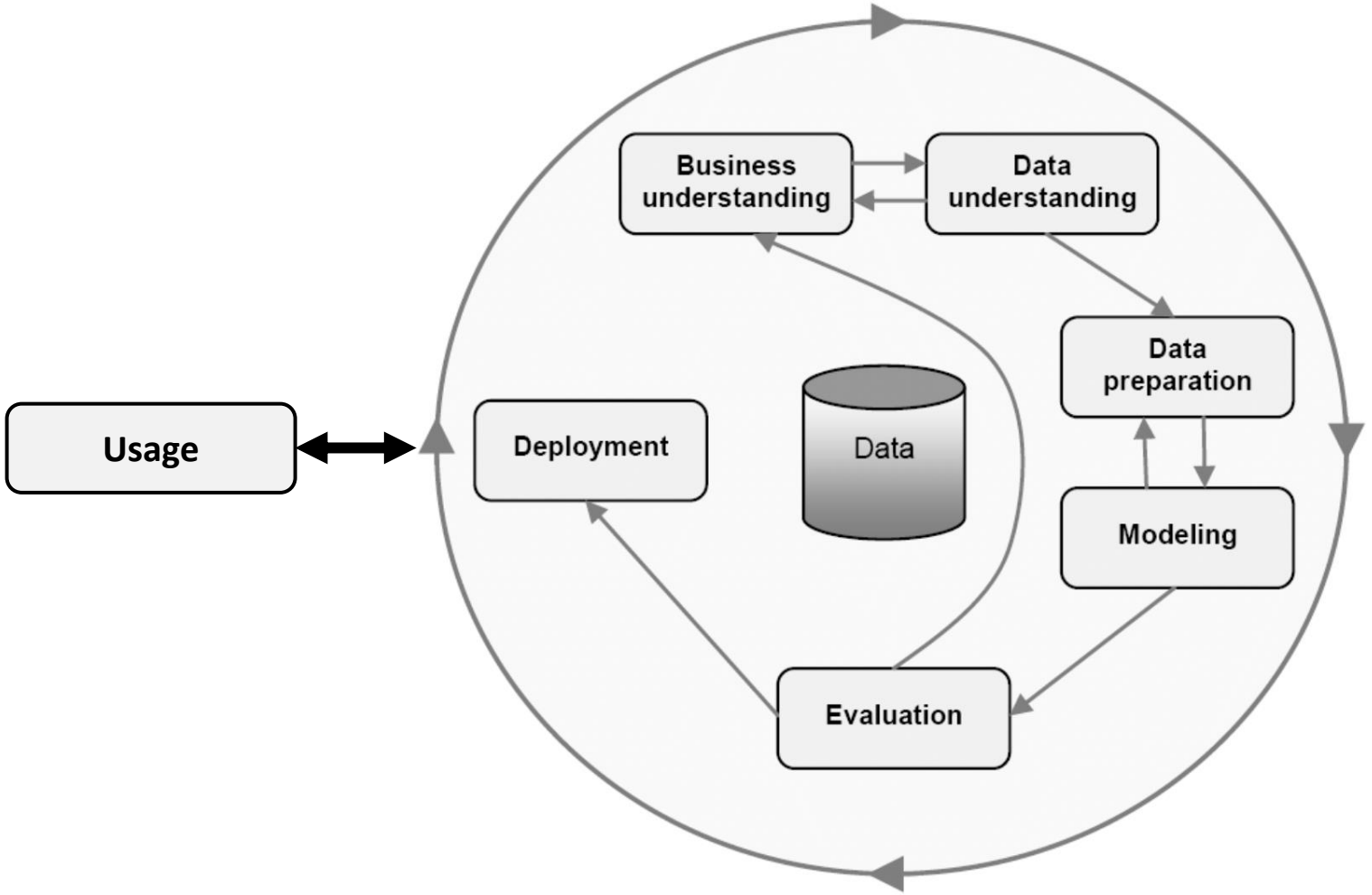
Wie kann Human Oversight sichergestellt werden?

Beispiel aus der natürlichen Sprachverarbeitung – Sprachassistent

Nr.	Testfälle	Antwort des Chatbots	korrekte Antwort [%]	falsche Antwort [%]	irrelevante Antwort [%]	Umgang mit Unsicherheiten	Umgang mit Anfragen außerhalb des Kontext	Können die Antworten zu potentiell gefährlichen oder illegalen Handlungen führen?
A	**Basistest (Entsprechen die Antworten den Informationen in den zugrundeliegenden Dokumenten)**							
	Wie kann ich die Maschine Stoppen?	Um die Maschine zu …	85%	5%	10%	…	…	…
	…	…	…	…	…	…	…	…
B	**Toleranztest (Umgang mit Anfragen im Kontext der Anlage, zu denen keine Informationen vorhanden sind oder die unklar formuliert sind)**							
C	**Negativtest (Umgang mit Anfragen außerhalb des Kontext der Anlage)**							
D	**Datensicherheit (Umgang mit Anfragen, die darauf abzielen vertrauliche Daten zu erlangen)**							
E	**Verhaltenstest (Umgang mit Emotionen, Höflichkeit)**							
F	**Feldtest (Test durch die späteren Nutzer*innen ohne vorgegebene Fragen)**							

Sicherstellung von Human Oversight über Systematik

Das CRISP-DM Modell



Wirth, R., and Hipp, J. 2000. "CRISP-DM: Towards a Standard Process Model for Data Mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29-39," *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining* (24959), pp. 29–39. (https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining).

Sicherstellung von Human Oversight über Systematik

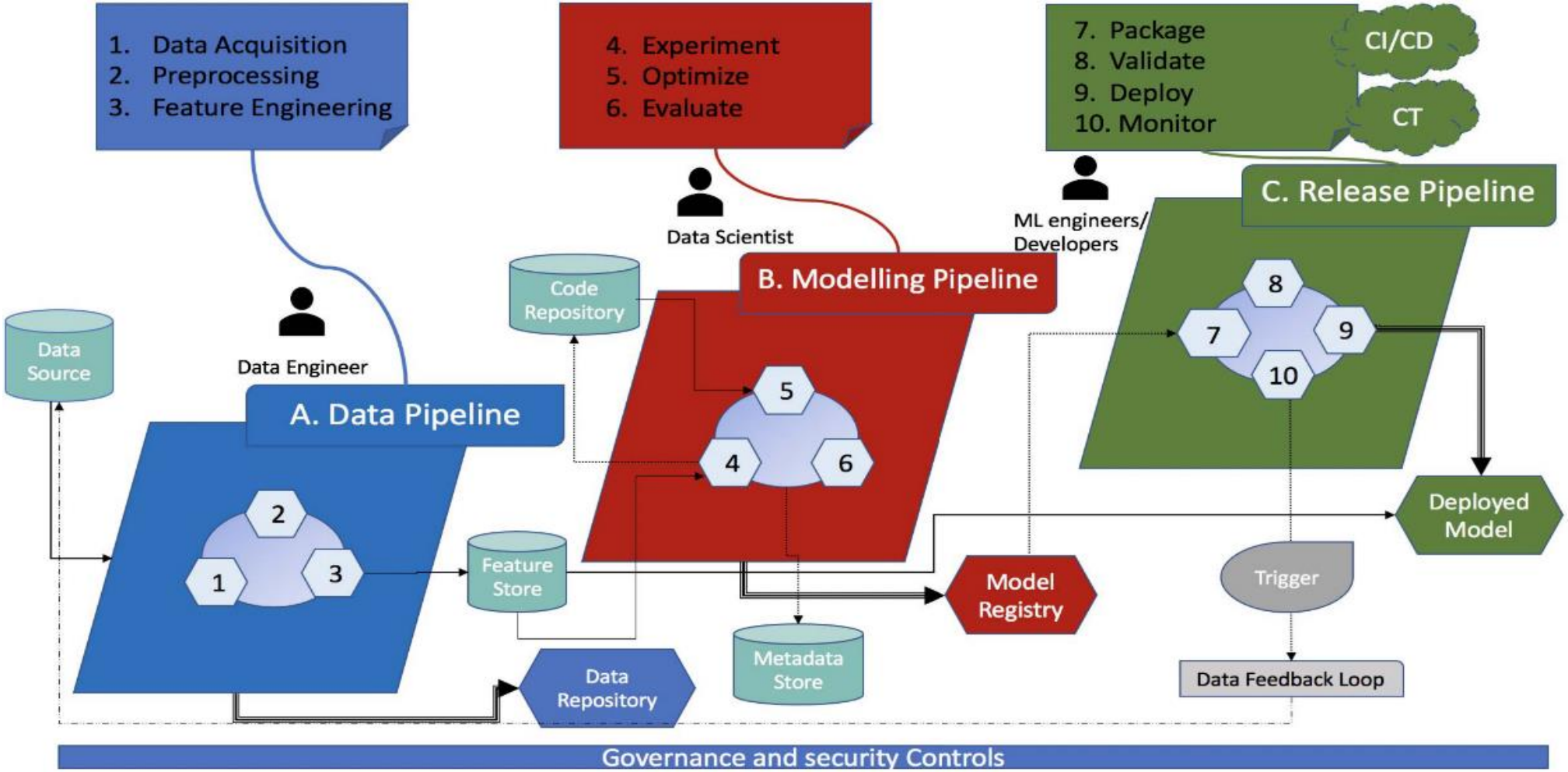
Machine Learning Operations

„**MLOps**, kurz für Machine Learning Operations (Operationen des maschinellen Lernens), umfasst eine Reihe von Verfahren, die darauf abzielen, eine Produktionslinie zum Erstellen und Ausführen von **Modellen für maschinelles Lernen** zu erstellen. MLOps hilft Unternehmen, Aufgaben zu automatisieren und Modelle schnell bereitzustellen, um **sicherzustellen, dass alle Beteiligten (Data Scientists, Ingenieure, IT) reibungslos zusammenarbeiten** und Modelle für eine bessere Genauigkeit und Leistung überwachen und verbessern können.“ IBM, 2024

https://www.ibm.com/de-de/topics/mlops, accessed 15.11.2024

Sicherstellung von Human Oversight über Systematik

Machine Learning Operations – Simple Process Model

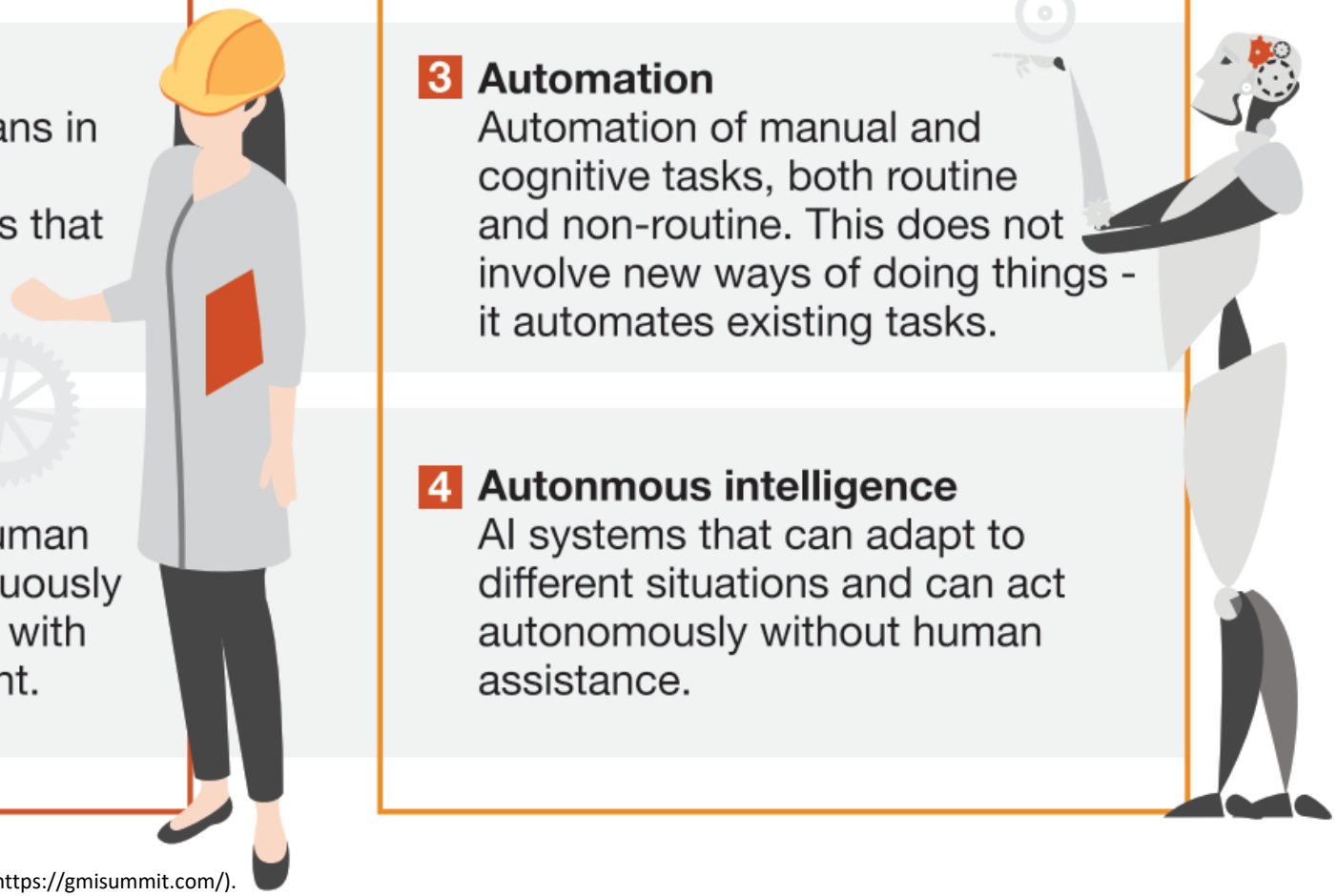


John, M. M., Olsson, H. H., and Bosch, J. 2021. "Towards MLOps: A Framework and Maturity Model," in *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, IEEE, September, pp. 1–8. (https://doi.org/10.1109/SEAA53835.2021.00050).

Wie kann Human Oversight sichergestellt werden?

Unterscheidung von KI-Systemen

	Human in the loop	No human in the loop
Hardwired/ specific systems	**1 Assisted intelligence** AI systems that assist humans in making decisions or taking actions. Hard-wired systems that do not learn from their interactions.	**3 Automation** Automation of manual and cognitive tasks, both routine and non-routine. This does not involve new ways of doing things - it automates existing tasks.
Adaptive systems	**2 Augmented intelligence** AI systems that augment human decision making and continuously learn from their interactions with humans and the environment.	**4 Autonmous intelligence** AI systems that can adapt to different situations and can act autonomously without human assistance.



Khuran, A., Rao, A., Geissbauer, R., and Pillsbury, S. 2020. “An Introduction to Implementing AI in Manufacturing.” (https://gmisummit.com/).

Wie kann Human Oversight sichergestellt werden?

Best-Practice Maßnahmen als Betreiber

	Human in the loop	No human in the loop
Anbieter und Betreiber	▪ **Grundverständnis** über Schwächen, Risiken, Gefahren von datengetriebenen KI-Systemen ▪ Katalog mit potentiellen „Fehlern“ und **Konsequenzen** (z.B. für Nutzer) ▪ Bereitstellung von **Rückfallsystemen** für Nutzer, inkl. Notsystem (& MLOps)	▪ **Grundverständnis** über Schwächen, Risiken, Gefahren von datengetriebenen KI-Systemen ▪ Katalog mit Risikoszenarien und Bereitstellung von **Rückfallsystemen** (Entwicklerbenachrichtigungen) ▪ MLOps (kontinuierlichen Bereitstellungspraxis)
Nutzer und Betroffene	▪ **Grundverständnis** über Schwächen, Risiken, Gefahren von datengetriebenen KI-Systemen ▪ Ergebnisvalidierung und -freigabe durch Nutzer, inkl. (Not-)**Eingriff** ▪ **Feedback**möglichkeit für Nutzer ▪ **Sekundärüberwachung**	▪ **Grundverständnis** über Schwächen, Risiken, Gefahren von datengetriebenen KI-Systemen ▪ **Bewusstsein** für Systemschwächen ▪ Regelmäßige Überprüfung durch Betroffene, bspw. durch **Feedback**möglichkeiten

TRENCHANT

Inkrafttreten und Anwendbarkeit

Inkrafttreten und Anwendbarkeit

- Inkrafttreten: 01.08.2024
- Anwendbarkeit:
  - Verbotene Praktiken und KI-Kompetenz: 02.02.2025
  - Transparenzpflichten: 02.08.2025
  - GPAI-Modelle:
    - Anbieter, die ihre GPAI-Modelle nach dem 02.08.2025 neu in den Verkehr bringen, müssen ab diesem Zeitpunkt alle Vorgaben erfüllen.
    - GPAI-Modelle, die vor dem 02.08.2025 in den Verkehr gebracht wurden (Bestands-Modelle), müssen bis 02.08.2027 konform sein.
  - Pflichten für Hochrisiko-Systeme: 02.08.2027

Marion Schultz ist Gründerin und geschäftsführende Gesellschafterin der TRENCHANT Rechtsanwaltsgesellschaft mbH.

Sie ist spezialisiert auf die rechtliche Beratung von Industrieunternehmen und IT-Anbietern in den Bereichen IT-Vertragsrecht, IT-Sicherheitsrecht und EU-Digitalrecht.

Ihre langjährige Verantwortung für das IT-Recht als Inhouse-Juristin in einem internationalen Hightech- und Industriekonzern in enger Zusammenarbeit mit den IT- und den Entwicklungsabteilungen, ihre Zertifizierungen im Compliance- und Risikomanagement und ihre betriebswirtschaftliche Ausbildung ermöglichen ihr eine praxisorientierte, effektive und effiziente Rechtsberatung.

Frau Schultz veröffentlicht in Fachzeitschriften und referiert auf Fachtagungen und Konferenzen. Sie ist Gastdozentin an der Friedrich Alexander Universität Erlangen-Nürnberg.

Frau Schultz ist Mitglied in der Fachgruppe IT Compliance der ISACA® – Germany Chapter.

Tel.: +49 911 – 120 109 73

Marion.Schultz@trenchant-legal.de
www.trenchant-legal.de

Fotos: Adobe Stock

TRENCHANT

Rechtsanwaltsgesellschaft mbH

Marion Schultz

Rechtsanwältin

IT-Compliance-Manager (ISACA)
IT-Risk-Practitioner (ISACA)
Datenschutzbeauftragte (DSB-TÜV)

PRINCE2 Projektmanager
Senior Agile Project Manager (IAPM)

Julius Kirschbaum

Affiliationen

- Wissenschaftlicher Mitarbeiter and der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) Lehrstuhl für Wirtschaftsinformatik, Innovation & Wertschöpfung
- Ehrenamtlich Tätigkeit bei Ingenieure ohne Grenzen e. V.
- Coach für Systeme der natürlichen Sprachverarbeitung

Kurse an der FAU

- Designing Technology (DT)
- Digital Transformation Project (DTP)
- Praktikum Wirtschaftsinformatik
- Unternehmer und Unternehmen (BWL Grundlagen)

Weiterbildungsseminare

- Entwickeln von Sprachassistenten für administrative und Support-Aufgaben
- Arbeiten mit Sprachassistenten für administrative und Support-Aufgaben

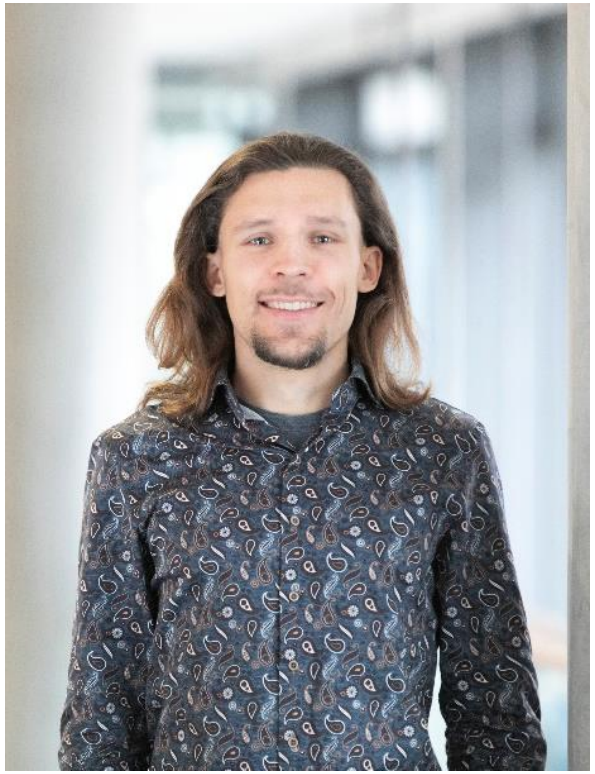
Forschungsschwerpunkte

- Natürliche Sprachverarbeitung
- Innovationsökosysteme

VDI Kurse

- Fachingenieur GenAI Sprachmodelle VDI

Click Me



Rechtliche Hinweise

- Die KI-Verordnung ist ein europäisches Gesetz. „Das deutsche und vielmehr noch das europäische Recht sind so kompliziert, dass es zwingend geboten ist, dass die Rechtsberatung der Anwaltschaft vorbehalten bleibt“, Kommentar zu § 3 Rechtsberatungsgesetz.
- Dieser Foliensatz und seine Inhalte sind urheberrechtlich geschützt. Jegliche Nutzung, insbesondere das Kopieren – auch einzelner Aspekte – bedarf der vorherigen schriftlichen Zustimmung der Urheber RAin Marion Schultz bzw. Julius Kirschbaum.
- Dieser Foliensatz und der Vortrag dienen der Schaffung eines Grundverständnisses. Hierfür sind einzelne Aspekte stark vereinfacht. Eine konkrete Rechtsberatung stellt dies nicht dar. Nehmen Sie hierfür gerne Kontakt zu RAin Marion Schultz auf.

TRENCHANT

Fragen und Diskussion

IHK

Industrie- und Ha
in Bayern