

ZWISCHEN HEILSPRECHEN UND SYSTEMISCHEM RISIKO:

ETHISCHE LEITLINIEN FÜR DIE VERANTWORTLICHE
KI-TRANSFORMATION

PROF. DR. ULRICH LENZ

Wie generative KI funktioniert

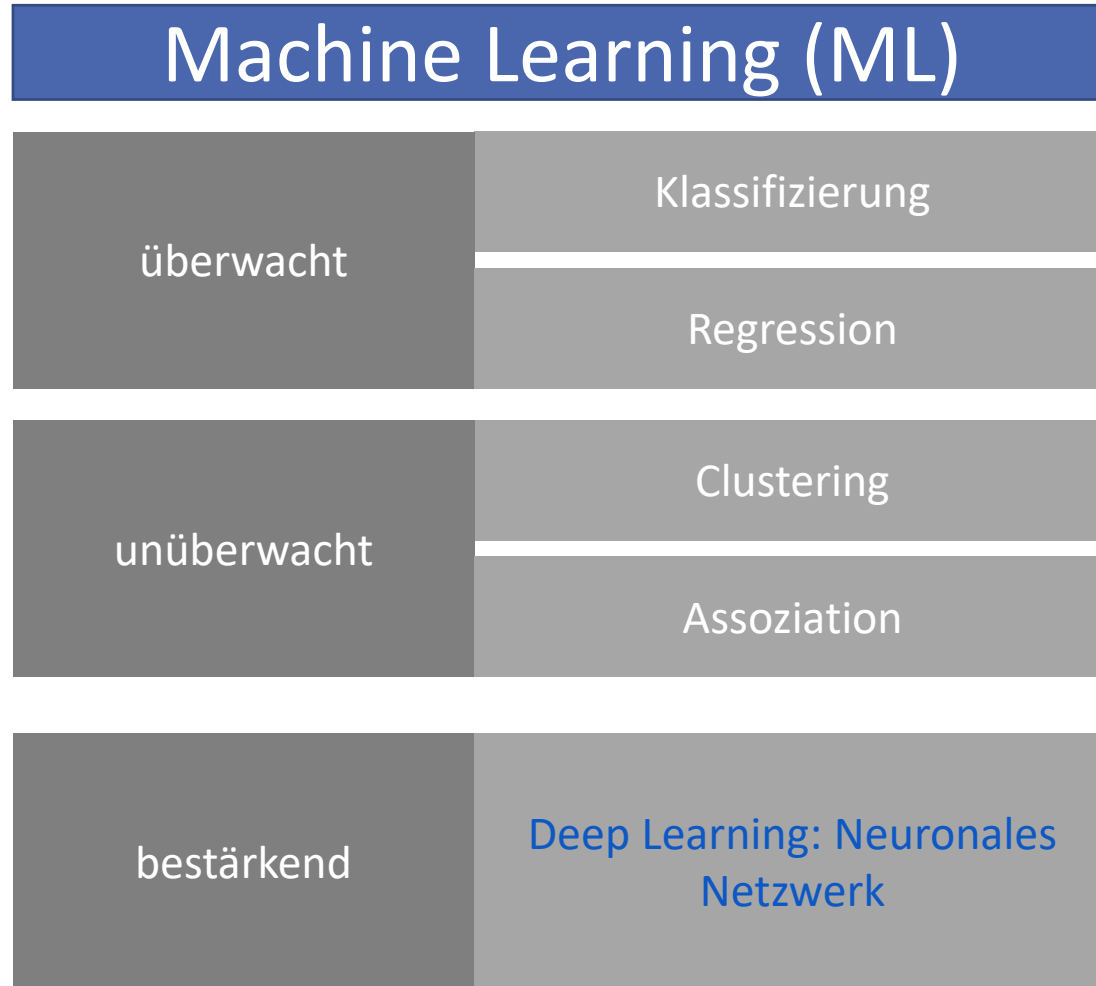
Spannungsfelder zwischen Heilsversprechen und existenziellen Risiken

European AI Act: Chance zu wirkungsvoller Regulierung, oder...?

Über den AI Act hinaus: Responsible AI

Themen für Evaluierende

KI ist nicht neu, aber das spezielle Lernverfahren ist es



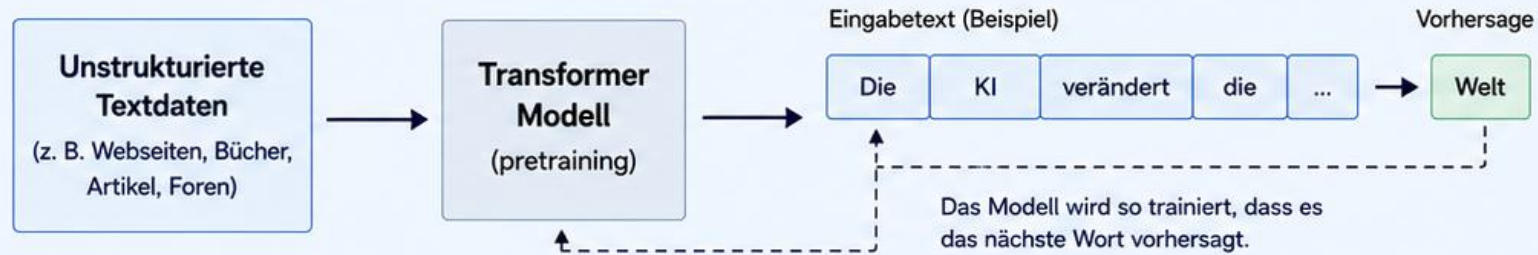
Trainingskosten bis zu ca. 200 Mio US\$

GPT: Generative Pretrained Transformer

G Generative (Pretraining)

Das Modell wird auf sehr großen Textmengen ohne manuelle Beschriftung vortrainiert.

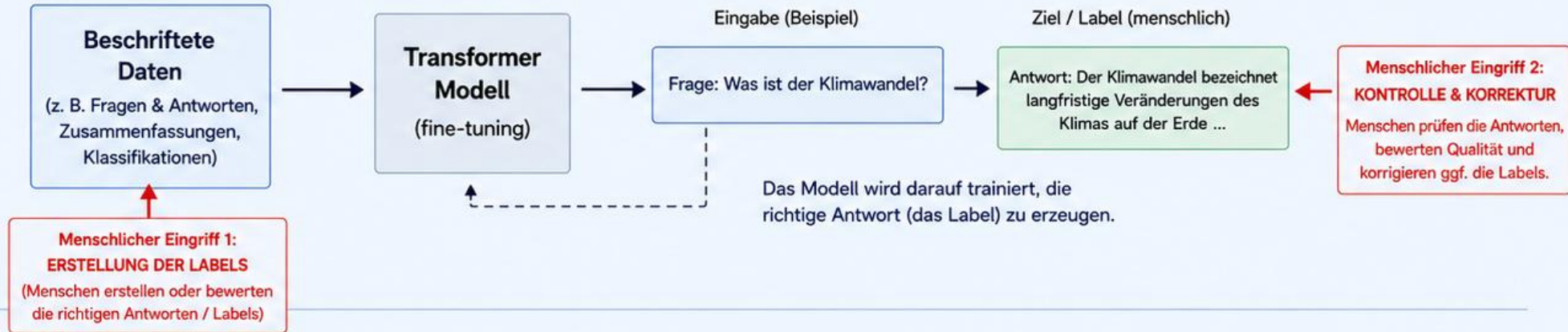
Ziel: Das Modell lernt die Struktur und Muster von Sprache, indem es das nächste Wort vorhersagt.



P Pretrained (Labelling / Fine-tuning)

Das vortrainierte Modell wird mit menschlich beschrifteten Daten für spezifische Aufgaben weitertrainiert.

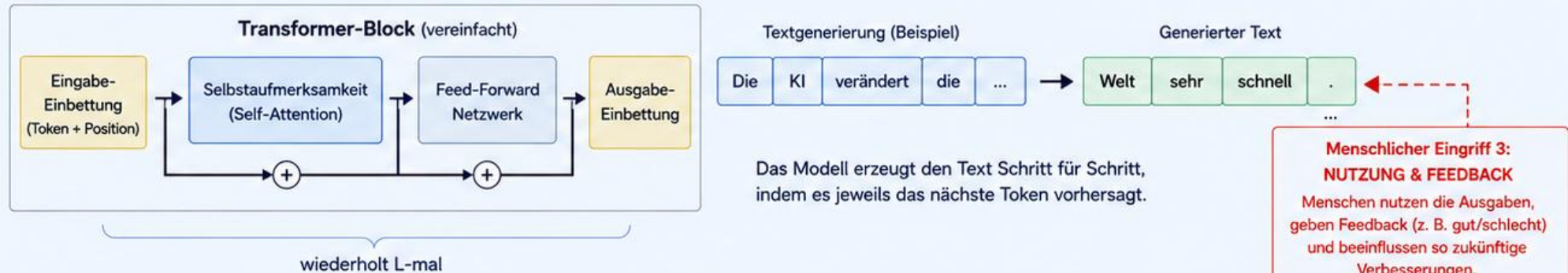
Ziel: Das Modell wird für konkrete Aufgaben leistungsfähig und zuverlässig.



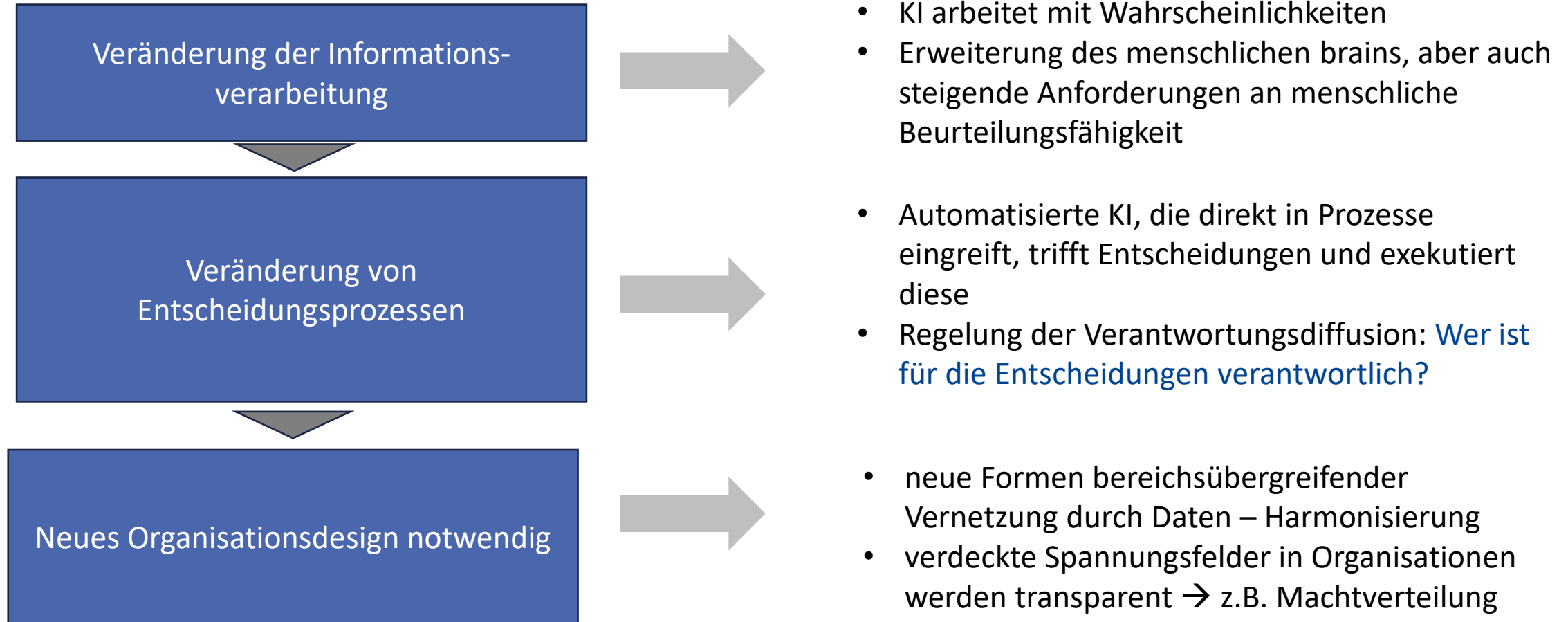
T Transformer (Architektur)

Die Transformer-Architektur ist das neuronale Netzwerk, das den Text verarbeitet und erzeugt.

Sie nutzt Selbstaufmerksamkeit (Self-Attention), um Zusammenhänge im Text zu erfassen.



KI ist nicht nur eine neue Anwendung, sondern verändert die Organisationslogik fundamental



Wie generative KI funktioniert

Spannungsfelder zwischen Heilsversprechen und existenziellen Risiken

European AI Act: Chance zu wirkungsvoller Regulierung, oder...?

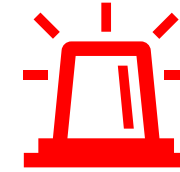
Über den AI Act hinaus: Responsible AI

Themen für Evaluierende

„Mythos“ – Eine KI von Anthropic



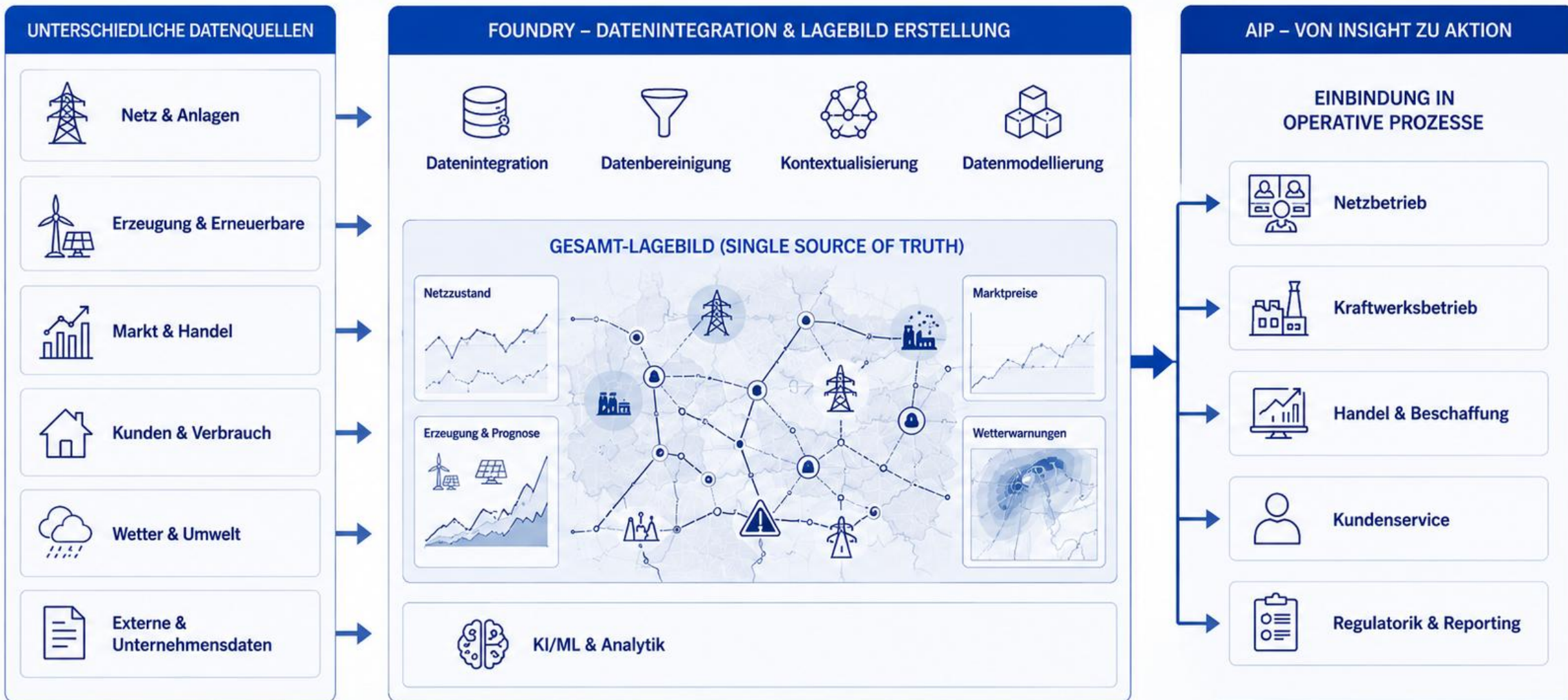
- Erweitertes chain-of-thought reasoning
- Verknüpfung von Informationen über sehr große Datenmengen
- Erweiterte Sicherheitsarchitektur → red-teaming-tests und sehr gute Eignung für Anwendungen in der kritischen Infrastruktur, im Rechtsbereich, in der Landesverteidigung



- Anwendung kann Sicherheitslücken in sensiblen Anwendungsbereichen finden: Banken, Energienetze, Sicherheitseinrichtungen
- „Ausbruch“ der KI in internem Test
- Aktuell Auslieferung an nur 11 Kunden ausschließlich in den USA

AIP Plattform von Palantir in Zusammenarbeit mit Foundry

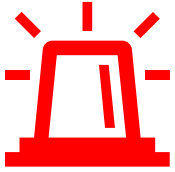
Von Daten zu Entscheidungen – Integriert in operative Prozesse der Energiewirtschaft





- **Single operational view** durch Integration strukturierter und unstrukturierter Daten aus unterschiedlichen Systemen
- **KI als operatives System**, das mit Workflows, Alarmierungs- und Steuerungssystemen verknüpft werden kann
- Starke **Governance-Funktionen**:
 - Auditierbarkeit
 - Protokollierung
 - Nachvollziehbarkeit der Datenherkunft
 - zentrale Sicherheitsrichtlinien lassen sich als Constitution formulieren, verbindlich für alle angebundenen Systeme

AI- Plattformen (z.B.) von Palantir: Deutliche Risiken



- Direkte Steuerung operativer Kern- und Sicherheitsprozesse
- Forward Deployed Engineers (FDE)
- Intransparente Tracking-Funktionen durch den Anbieter

Daraus folgen grundsätzliche ethische Probleme und Fragen:

- Lebensbedrohliche Fehlentscheidungen durch die KI: Bekämpfung ziviler Ziele aufgrund veralteter oder manipulierter Daten
- Autonome KI-Kampfroboter
- Vendor lock in
- Postulat der menschlichen Letztentscheidung?

The jagged frontier-Studie*: Neue Formen der Zusammenarbeit zwischen Mensch und KI hat auch eine ethische Komponente

Bearbeitung komplexer, wissensintensiver Aufgaben in einem Laborexperiment, n=758 BCG Berater

Kein Zugang zu KI

Zugang zu KI

Zugang zu KI und Prompt Guide

Innovative Aufgabe: Produktentwicklung
→ innerhalb der „Grenze“

Analytische Aufgabe: Komplexe Problemlösung
→ außerhalb der „Grenze“

Fabrizio Dell’Acqua,^a Edward McFowland III,^a Ethan Mollick, Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality. ORGANIZATION SCIENCE Vol. 37, No. 2, March–April 2026, pp. 403–423

The Jagged-Frontier-Studie: Erkenntnisse

„Grenze“: Aufgaben, die aus Sicht der Berater gleichen Schwierigkeitsgrad haben...

- werden, auch wenn sie scheinbar **innerhalb der „Grenze“** liegen, von der KI mit unterschiedlicher Qualität gelöst
- mit KI-Unterstützung **12% mehr Aufgaben, 25% schneller, höhere Qualität**
- **komplexe** Beratungsaufgaben wurden von den Teams mit KI-Unterstützung **20% schlechter** erledigt.

The Jagged-Frontier-Studie: Ethische Herausforderungen

Organisationssicht

- KI-Einsatz kann menschliche Arbeit verdrängen
- Automatisierung von Abläufen erhöht die Leistungsfähigkeit von Organisationen, ermöglicht beschleunigte Genehmigungs- und Bescheidverfahren und hat das Potenzial, die Kund*innen – Zufriedenheit zu erhöhen

Perspektive der Mitarbeitenden

- KI-Einsatz kann menschliche Arbeit verbessern
- KI-Einsatz kann Inklusion verbessern
- KI-Einsatz kann den Menschen aufwerten und Arbeitszufriedenheit durch bessere Arbeitsqualität steigern



Ethisches Dilemma: Automatisierung mit ggf. Verdrängung von Arbeitskraft versus Stärkung der Mitarbeitenden durch Human Centered AI

Verallgemeinerung: Dilemmata der Technikethik

Betrachtungs-
gegenstand

Ethische Fragen der Entwicklung von Technologien,
der Nutzung und der Technikfolgen

Dilemmata

- Datenverfügbarkeit <-> Privatsphäre
- Regulierung <-> Innovation
- Automatisierung <-> menschliche Arbeit

Spezifizierung: KI-Assistenten als technikethisches Problem

Betrachtungs-
gegenstand

KI-Assistenten führen aufgrund vorgegebener Entscheidungsregeln eigenständig komplette Arbeitsprozesse durch.

Beschreibung
des Dilemmas

Datenschutz / Privatsphäre versus Arbeitserleichterung:

Bereits ein „harmloser“ Agent, der einen priorisierten Tages-Arbeitsplan erstellen soll, benötigt einen vollen Zugriff auf alle Emails, Aufgaben und Kalendereinträge der Mitarbeitenden, die den Assistenten einsetzen wollen oder sollen.

KI-Assistenten für anspruchsvolle medizinethische Aufgaben – Erledigung wird das Dilemma dramatisch!

Neue Herausforderungen für Technik-Ethik

Oder: KI verlässt die KI – Bubble*

- ➔ **Verantwortungsvolle KI hält nicht Schritt mit der Entwicklung von KI-Innovationen:** Ernste KI-Vorfälle 2025 = 362, 2024 = 233
- ➔ **Medizinethik:** Bis zu 83% weniger menschlicher Dokumentationsaufwand, aber methodische Probleme wissenschafts-ethischer Studien, da kaum echte Patientendaten verwendet
- ➔ **Zunehmende Systemkosten:** Chipentwicklung, Lieferkettenrisiken für Rohstoffe für die Chips, Energiekosten für die Rechenzentren:
 - Training der AI Grok emittierte 73 T Tonnen CO2
 - KI Rechenzentren der USA verbrauchen 30 GW Strom → äquivalent zum Energieverbrauch des Staates New York

[ai_index_report_2026.pdf](#)

Lösungsansätze für die ethischen Herausforderungen



Regulierung: European AI Act



Technische Maßnahmen in der Entwicklungsphase eines KI Algorithmus



Nicht-technische Maßnahmen: Selbstverpflichtungen der KI-Anbieter, organisationsinterne Governance-Konzepte, Aufbau eines wirkungsvollen KI-Kompetenzmanagements für alle Mitarbeitenden

Wie generative KI funktioniert

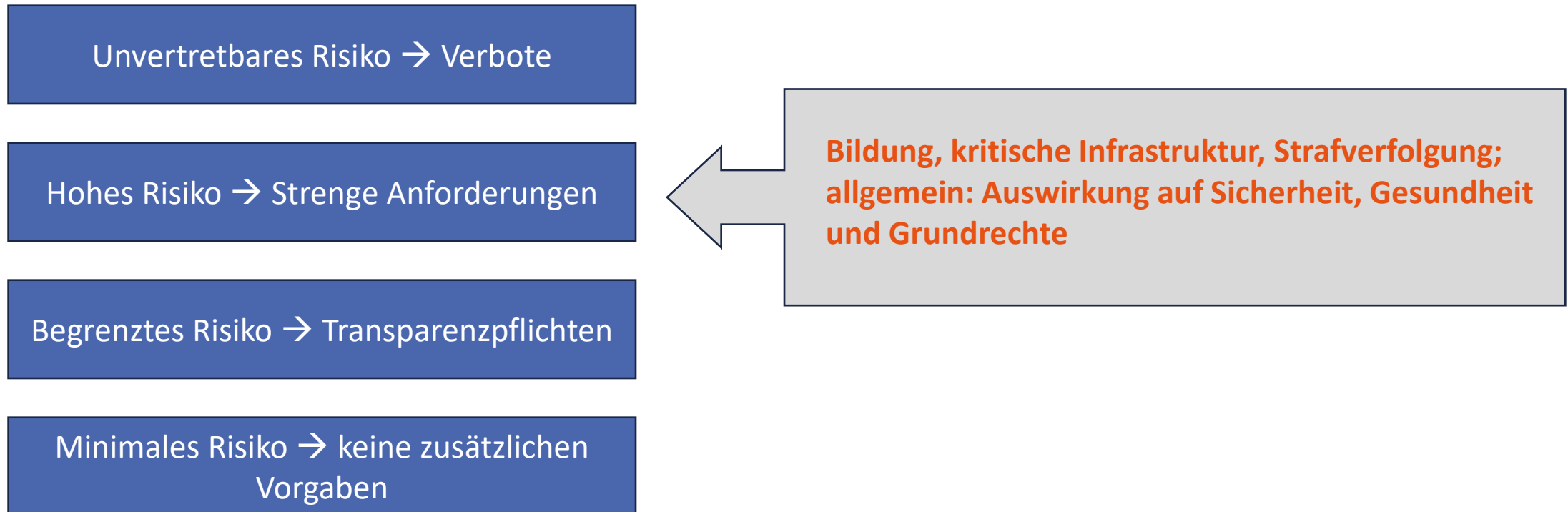
Spannungsfelder zwischen Heilsversprechen und existenziellen Risiken

European AI Act: Chance zu wirkungsvoller Regulierung, oder...?

Über den AI Act hinaus: Responsible AI

Themen für Evaluierende

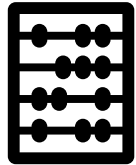
Risikobasierter Ansatz: Kopplung des Regulierungs – Umfangs an das potenzielle Risiko der Anwendung



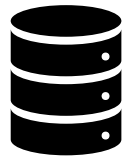
European AI Act: Ein sehr knapper Überblick

- Gilt auch für organisationsintern entwickelte KI-Anwendungen
→ „in Verkehr bringen“
- Weitergehende Regelungen für GPAI → Art. 53 EU AI Act
- Hochrisiko-Anwendungen müssen zertifiziert werden und in einer Datenbank registriert sein
- **Für Entwicklung und Betrieb von Hochrisiko-KI beachten:**
 - Sicherheitstechnologie
 - Dokumentation
 - Datenschutz
 - Transparenzpflichten
 - Konformität
 - Menschliche Aufsicht

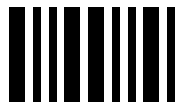
In-Verkehr-bringen von KI erfordert vorherige Qualitäts- und Konformitätssicherung



Für Hochrisiko-KI: Die neu entwickelte – oder grundlegend veränderte - KI muss den EU-Zertifizierungsprozess durchlaufen, u.a. Sicherstellung der menschlichen Letztkontrolle der Anwendung



Registrierung der KI in einer EU - Datenbasis



Unterzeichnung einer Konformitätserklärung und CE-Kennzeichnung

Wirkungsvolle Regulierung oder Bürokratie-Monster?

Für wirkungsvolle Regulierung spricht...

- Rechtssicherheit
- Planungssicherheit insbesondere durch den risikobasierten Ansatz
- Verfahrenssicherheit: Implementierungshinweise sind verfügbar
- Gültigkeit **alle** Anbieter, die in der EU KI in Verkehr bringen wollen, auch für Eigenentwicklungen in Organisationen bzw. Behörden

Abgesehen davon, dass es sich um rechtlich verbindliche Regelungen handelt, spricht für Bürokratiemonster...

- Umfangreiches Gesetzeswerk, das mehrere Handreichungen / Implementierungshinweise erfordert
- Erfordert administrativen Aufwand und ggf. zusätzliches Personal für Entwicklung, Implementierung und Überwachung
- Läuft der technologischen (KI-) Entwicklung hinterher

Wie generative KI funktioniert

Spannungsfelder zwischen Heilsversprechen und existenziellen Risiken

European AI Act: Chance zu wirkungsvoller Regulierung, oder...?

Über den AI Act hinaus: Responsible AI

Themen für Evaluierende

KI-Ethik – Anforderungen umsetzen:

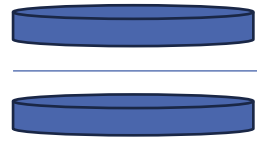
Technische Mitüberwachung durch die KI



Ethische Grundsätze in konkrete Anweisungen überführen



Constitutional AI prüft intern vor Ausgabe der Antwort, ob Regeln eingehalten wurden



Policy Layer einfügen



KI prüft Eingaben und Ausgaben:
Zugriffsteuerung,
Compliance-Prüfung,
Inhaltsfilterung



Prompt Injection



Prüfmechanismen gegen manipulative Eingaben



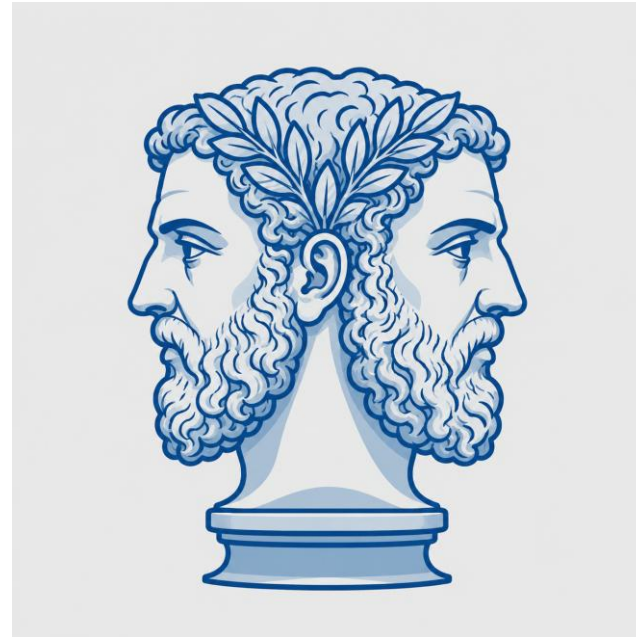
Technische Begrenzungen (Guardrails): Regelbasierte Prüfungen und Validierung der Ausgaben

Nicht-technische Maßnahmen: Die Sensibilität gegenüber den potenziell negativen Folgen der KI ist deutlich angestiegen.



GPAI oderspezialisierte
Eigenentwicklung?

→ Ethik muss früh in
den Entwicklungsprozess
eingebaut werden



Wer reguliert die
„Constitution“?

Anbieterspezifische KI
Verfassung nicht nur aus
ethischen Gründen motiviert,
sondern auch Geschäftsmodell

Technologiesprünge und neue Herausforderungen – Oder: KI verlässt die KI – Bubble*



KI-Souveränität als politische und wirtschaftliche Herausforderung angesichts steigender Abhängigkeiten



Vertrauen in den Staat, die KI regulieren zu können, sinkt.

- ➔ Menschen sind angesichts der Fähigkeiten – Sprünge der KI geneigt, der KI freiwillig Macht einzuräumen.* Datensicherheit wird dann als weniger wichtig angesehen.
- ➔ Die Lücke zwischen Integration ethischer Grundsätze in KI und dem Druck auf unlimitierte Datenverfügbarkeit geht zunächst noch auseinander
- ➔ A race you cannot win: Staatliche Regulierung ist immer langsamer als die technologische Entwicklung. Für eine wirksame KI-Ethik sind weitere technische und nicht-technische Maßnahmen notwendig**

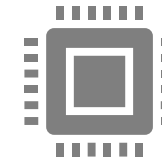
*Lenz, U. (2023). Machtverschiebung durch generative Künstliche Intelligenz - und die Konsequenzen für das Transformationsmanagement (S. 291-307). In: Geramanis, O., Hutmacher, S., Walser, L. (Hrsg.) Organisationale Machtbeziehungen im Wandel - Führung zwischen Zustimmung und Zwang: Springer

**Lenz, U. (2024). Generative Künstliche Intelligenz: Chancen und Stolpersteine für die Entwicklung von Organisationen. Organisationsberatung, Supervision, Coaching (OSC): Springer Fachmedien

KI-Ethik Anforderungen umsetzen: Responsible AI geht nur *mit* dem Menschen



- Situationsspezifische Kontext – Sensitivität
- Kompetenzentwicklung der Mitarbeitenden → Interventionsfähigkeit weiterentwickeln



- KI kann schneller skalieren → häufig sehr umfangreiche Validierungen notwendig
- Verifikation einer Ausgabe nach der anderen macht Geschwindigkeitsvorteile zunichte



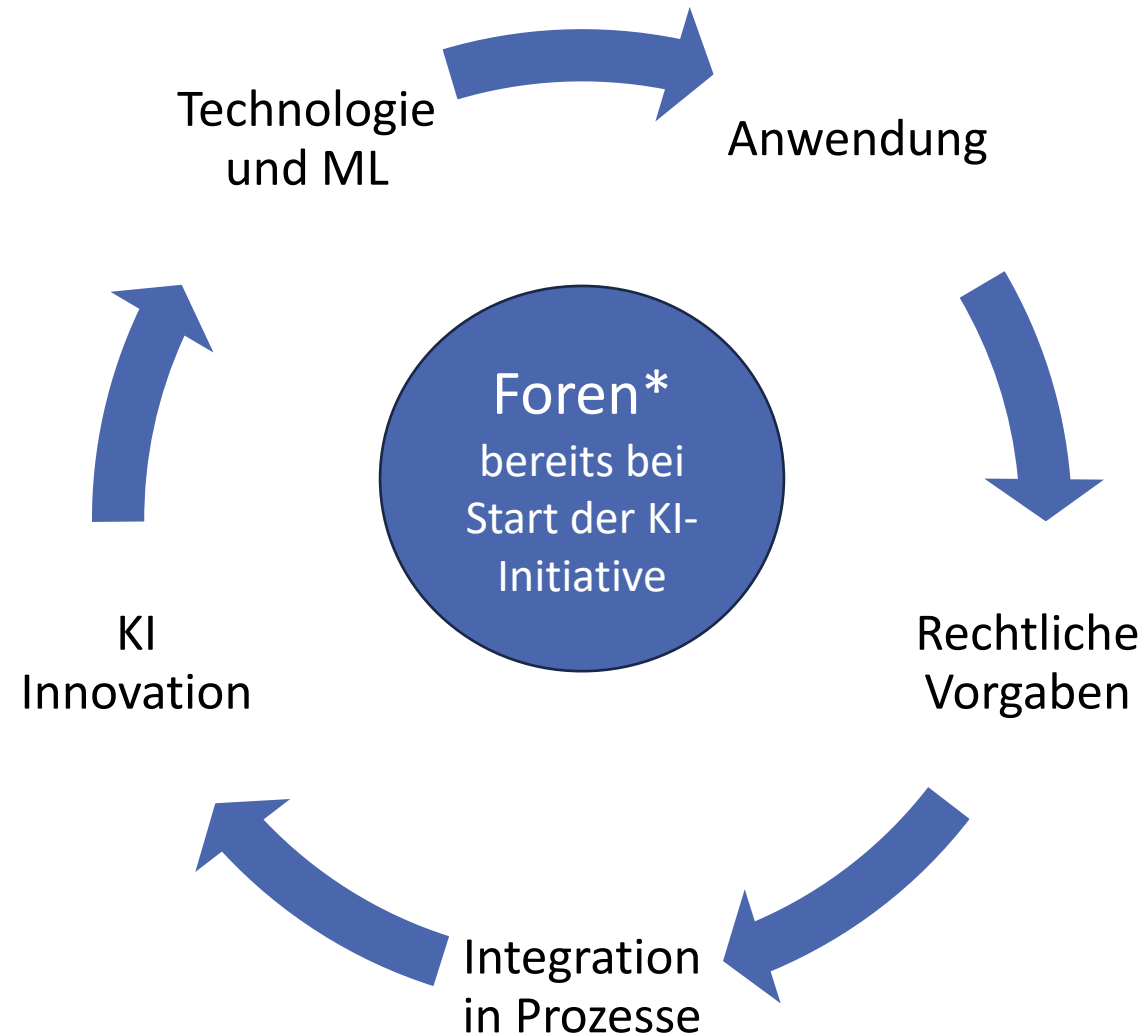
Umsetzung des Prinzips der menschlichen Letztverantwortung

KI Ethik praktisch umsetzen:

KI – Kompetenz auf- und ausbauen

- ➔ Der Mensch ist und bleibt unverzichtbar für die Beurteilung des KI-Outputs und für das Weiterlernen der KI
 - ➔ Die höchste Kompetenz zur Prüfung des Outputs liegt bei den Operations → dort wird durch KI-Kompetenzentwicklung Akzeptanz geschaffen
 - ➔ KI kann nur dann resilient und effektiv betrieben werden, wenn Mitarbeitende umfangreiche KI-Kompetenz haben
- **Entwicklung von KI-Kompetenz ist eine strategische Aufgabe**
 - **... die weit über ein „Training der Anwendung“ hinausgeht**

KI-Ethik umsetzen: Themen der KI – Kompetenz



*Austausch-Formate zu (z.B.)

- Anwendungserfahrungen
- Evaluierung
- Führung & Organisationsentwicklung

Wie generative KI funktioniert

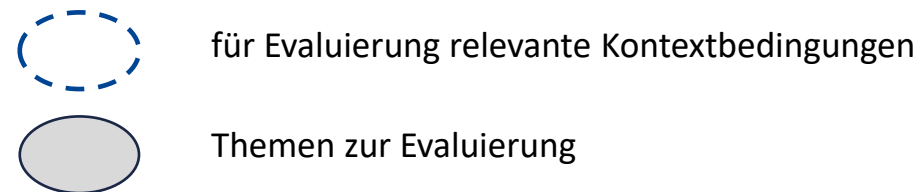
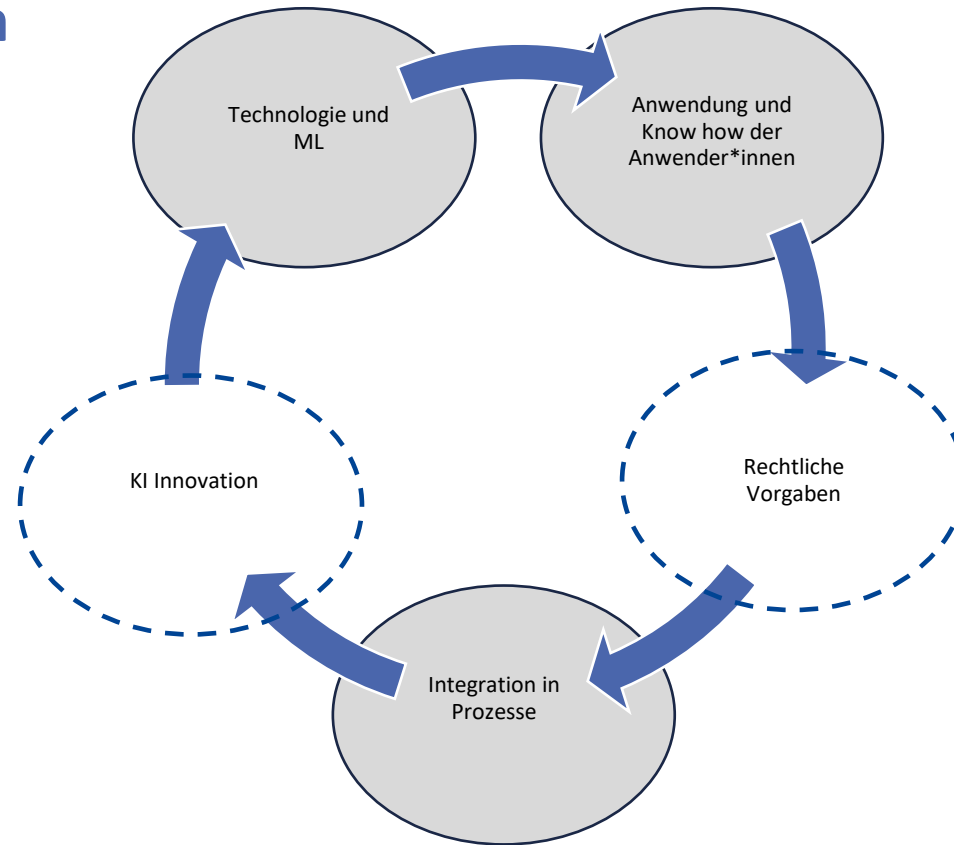
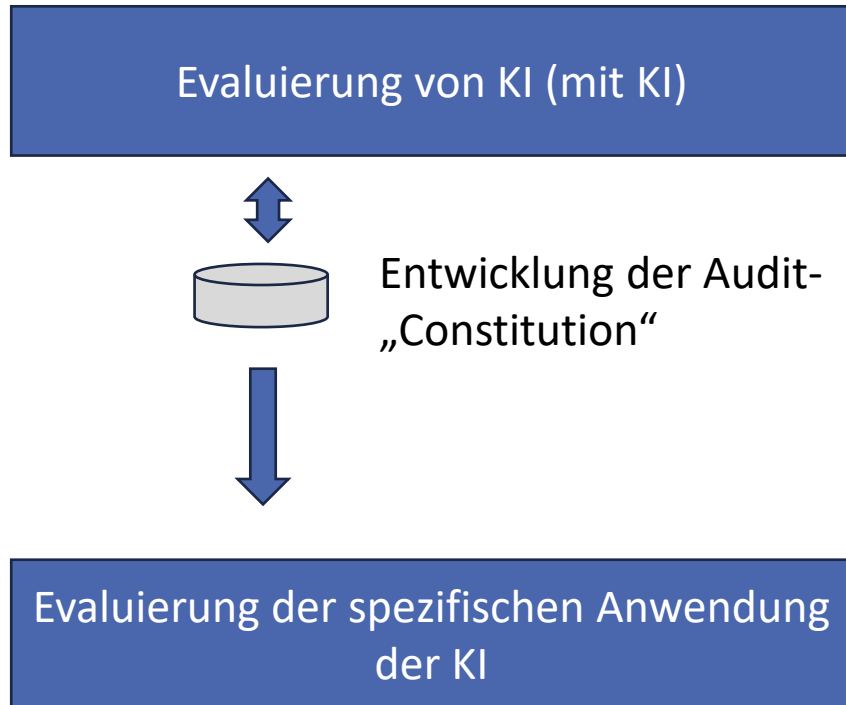
Spannungsfelder zwischen Heilsversprechen und existenziellen Risiken

European AI Act: Chance zu wirkungsvoller Regulierung, oder...?

Über den AI Act hinaus: Responsible AI

Themen für Evaluierende

Impuls zur Evaluierungs-Diskussion





Mail: Ulrich.Lenz@outlook.com
Web: <https://lenz-advisoryservices.com/>
LinkedIn: <https://www.linkedin.com/in/ulrich-lenz/>
Tel.: +49 151 56120796

Services

Consulting und Transformations- begleitung

Individuelles, fachliches Sparring für Projektverantwortliche in Digitalisierungs-Initiativen auf der Erfahrungs- Grundlage der Leitung großer, komplexer Digitalisierungs – Initiativen und ca. 25 Jahre Führungserfahrung auf Bereichsleiter-Ebene

Leistungsspektrum

Begleitung der Strategieentwicklung und –überprüfung mit Ableitung eines sinnvollen KI-Einsatzes

Setup des Projekts, Vorbereitung und Moderation von Kick-off-Workshops

Unterstützung des Projektmanagements, Moderation von Retros und Reviews, Teamentwicklung für das Projektmanagement, Aufbau eines strategischen Stakeholder-Managements

Unterstützung bei der Planung und Steuerung des Transformationsprozesses zur Implementierung digitaler Innovationen einschließlich Beratung zu sinnvollen Maßnahmen der Organisationsentwicklung und der Change Kommunikation

Konzeption und Moderation von Workshops in allen Phasen eines Transformationsprozesses

Coaching

Executive Coaching

für C-level und Bereichsleitungen sowie Leiter*innen von Reorganisations- und Digitalisierungsprojekten

Teamcoaching

Teamentwicklung für Führungs- und Hochleistungsteams

Besonderheit: Anwendung wissenschaftlich fundierter therapeutischer Konzepte im Business Coaching

Vorträge und Workshops

zu allen zuvor genannten Themenbereichen, individuell konzipiert für Ihre spezifischen Fragestellungen und Anlässe