

Arbeitsplanung und Prävention bei der Gestaltung von KI-Systemen

– Aufgaben und Herausforderungen für Normung und Standardisierung

Dr.-Ing. Patricia Stock

Mainz, 18. April 2023



*KI hat größeres
Potenzial als die
Dampfmaschine.*

(McKinsey Global Institute 2018)

*Aber was ist eigentlich
KI (bzw. AI)?*

Definition “AI system” nach ISO/IEC FDIS 22989:2022

engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives

Definition “KI-System” nach dem AI Act der Europäischen Kommission

eine Software, die mit einer oder mehreren der in Anhang I aufgeführten Techniken und Konzepte entwickelt worden ist und im Hinblick auf eine Reihe von Zielen, die vom Menschen festgelegt werden, Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das Umfeld beeinflussen, mit dem sie interagieren;

Benannte Techniken und Konzepte in Anhang I: Ansätze zum Maschinellen Lernen, Logik- und wissensbasierte Ansätze, statistische Modelle, Bayes'sche Schätzung sowie Such- und Optimierungsmethoden

KI-Technologien

- Automatische Bilderkennung
- natürliche Sprache
- virtuelle Assistenten
- Roboter-basierte Prozessautomatisierung
- fortgeschrittenes maschinelles Lernen

*„Automation applied to an inefficient operation will magnify the inefficiency.“
(Bill Gates)*

Quelle: McKinsey Global Institute 2018

Voraussetzung für die Einführung und den Betrieb von KI-Technologien sind stabile Produktionssysteme, klar definierte Prozesse und Daten.

Gestaltung von KI-Systemen: Kognitiv-unterstützende Assistenzsysteme

Ziele für den Einsatz:

- Bereitstellung von Information über das Arbeitssystem
- Anleitung bei der Durchführung der Arbeitsaufgabe
- Ermöglichen von Lernprozessen
- Verbesserung der Zuverlässigkeit der Arbeitsperson
- Verbesserung der Zusammenarbeit
- ...

Zu klärende Fragen bei der Gestaltung:

- Aufgabenteilung?
- Benötigte Informationen? Datensicherheit?
- Übermittlung der Informationen?
- Ergonomische Gestaltung? Arbeitsschutz?
- Prozess- und Qualitätssicherung?
- Vermeidung von Ablenkung?
- Kontrolle? Verantwortung?
- Einführung? Überwindung von Hemmschwellen?
- ...

Künstliche Intelligenz: Relevante Themenfelder für den Arbeitsschutz

Schutz der Beschäftigten im Betrieb (Adressat Betreiber)

- Sicherheit und Gesundheit, physisch und psychisch im Betrieb

Sichere KI-Arbeitsmittel (Adressat Hersteller)

- Zuverlässigkeit und Vorhersagbarkeit, Risikoanalyse für Algorithmen
- Robustheit, Widerstandsfähigkeit der Systeme
- Datensicherheit, Datenqualität (Input) und Ergebnisqualität, Validität (Output)
- Transparenz, Erklärbarkeit, Systemvertrauen erzeugen, Übervertrauen vermeiden

Menschengerechte Arbeitsgestaltung

- Handlungs- und Entscheidungsspielräume
- Lern- und Kompetenzförderlichkeit
- Anforderungs- und Belastungsoptimierung
- Kommunikation, Kooperation, Interaktion
- Adaptivität, Individualisierbarkeit
- Fehlertoleranz

Organisationale Grundlagen

- Verantwortung bei Mensch- und KI-Schnittstellen definieren, Kongruenzprinzip von Befugnis und Verantwortung
- Partizipation und Feedbackschleifen realisieren
- Inadäquate Leistungskontrolle verhindern

Ethische und rechtliche Rahmenbedingungen

- Rechtsrahmen klären: Produktsicherheit, Betriebssicherheit, Haftung
- Datenschutz-, Persönlichkeits-, Anti-Diskriminierungsrechte

Normen, Standards, Zertifikate

- Normungsobjekte beschreiben: Entwicklungsprozess, Nutzungsprozess, Systemeigenschaften
- Normungsprozesse initiieren, internationale und europäische Normung
- Nutzen, Chancen, Risiken und Hindernisse bei möglichen Zertifizierungsprozessen

Quelle: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA); https://www.baua.de/DE/Angebote/Publicationen/Aktuell/1-2020.pdf?__blob=publicationFile&v=5

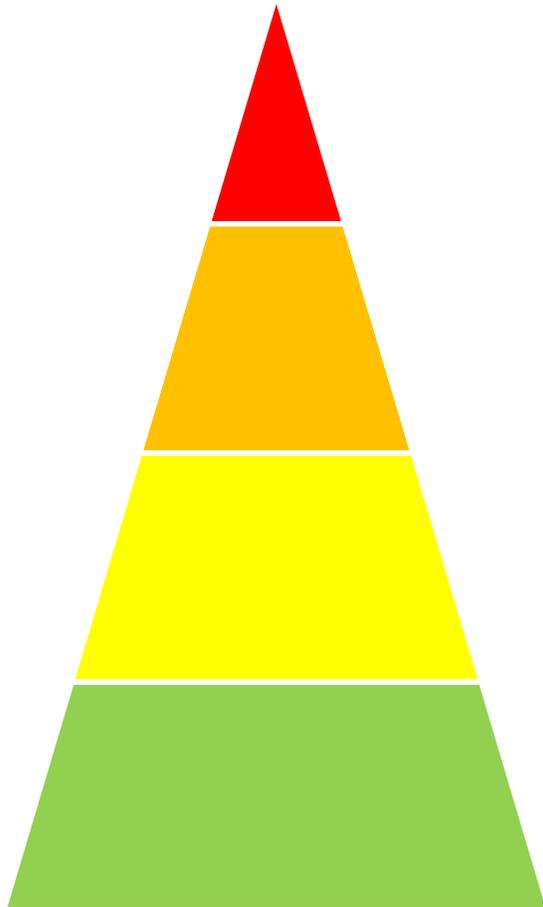
Regulation im Umgang mit KI durch die Europäische Union mit dem AI Act

Artificial Intelligence Act [COM(2021) 206 final]

Verordnung zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz)

- Regelung im Umgang mit KI in der Forschung und Wirtschaft
- Relevant für
 - Anbieter, die KI-Systeme in der EU in Verkehr bringen oder in Betrieb nehmen
 - Nutzer von KI-Systemen, die sich in der Union befinden
- Kategorisierung von KI-Systemen in vier Risikoklassen
→ Ableitung von Anforderungen für den Umgang mit KI-Systemen
- Ableitung von „Standardisation Requests“ für die Normung, u.a. bzgl. Risikomanagement-Systeme, Transparenz und menschlicher Aufsicht

Risikoklassen des AI Act der Europäischen Union



Unvertretbares Risiko z.B. unterschwellige Beeinflussung, Social Scoring	Verboten
Hohes Risiko z.B. Betrieb kritischer Infrastrukturen, Human Resource Management, Bildung, Strafverfolgung	Zulässig vorbehaltlich der Einhaltung der KI-Anforderung und der Ex-ante Konformitätsbewertung
Mäßiges Risiko z.B. Bots, Deep Fakes, Emotionserkennung	Zulässig unter der Voraussetzung von Informations-/Transparenzverpflichtungen
Minimales Risiko	Zulässig ohne Einschränkungen Verhaltenskodex vorgeschlagen

Die Grenzen sind oft fließend!

Die Normungsroadmap KI (Ausgabe 2) konkretisiert die Standardisierungsbedarfe

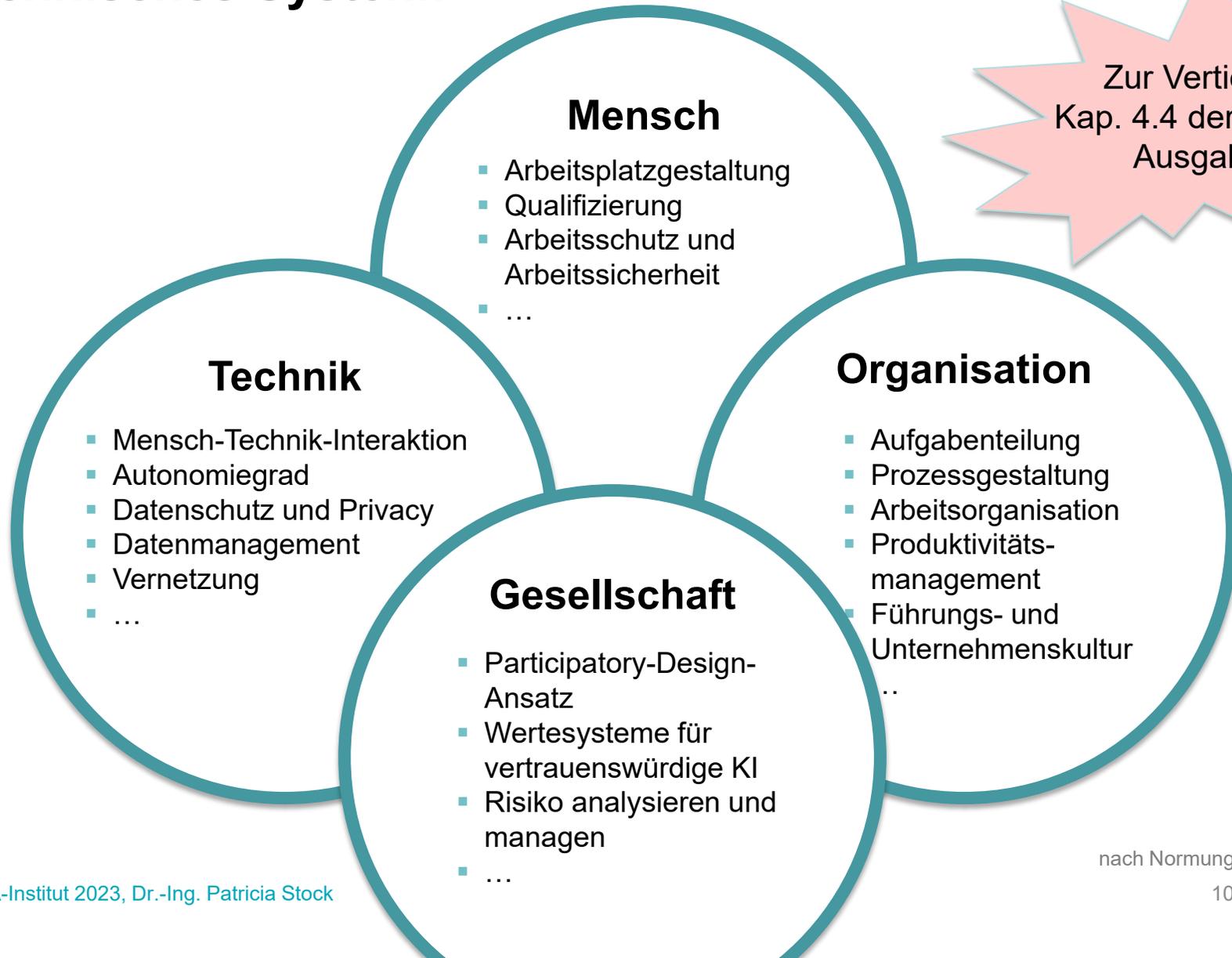


- Erschienen am 9.12.2022
- Download:
<https://www.dke.de/resource/blob/2008010/776dd87a4b9ec18d4ab295025ccbb722/nr-ki-deutsch---download-data.pdf>
- Adressierte Themen
 - Grundlagen
 - Sicherheit
 - Prüfung und Zertifizierung
 - **Soziotechnische Systeme**
 - Industrielle Automation
 - Mobilität
 - Medizin
 - Finanzdienstleistungen
 - Energie/Umwelt

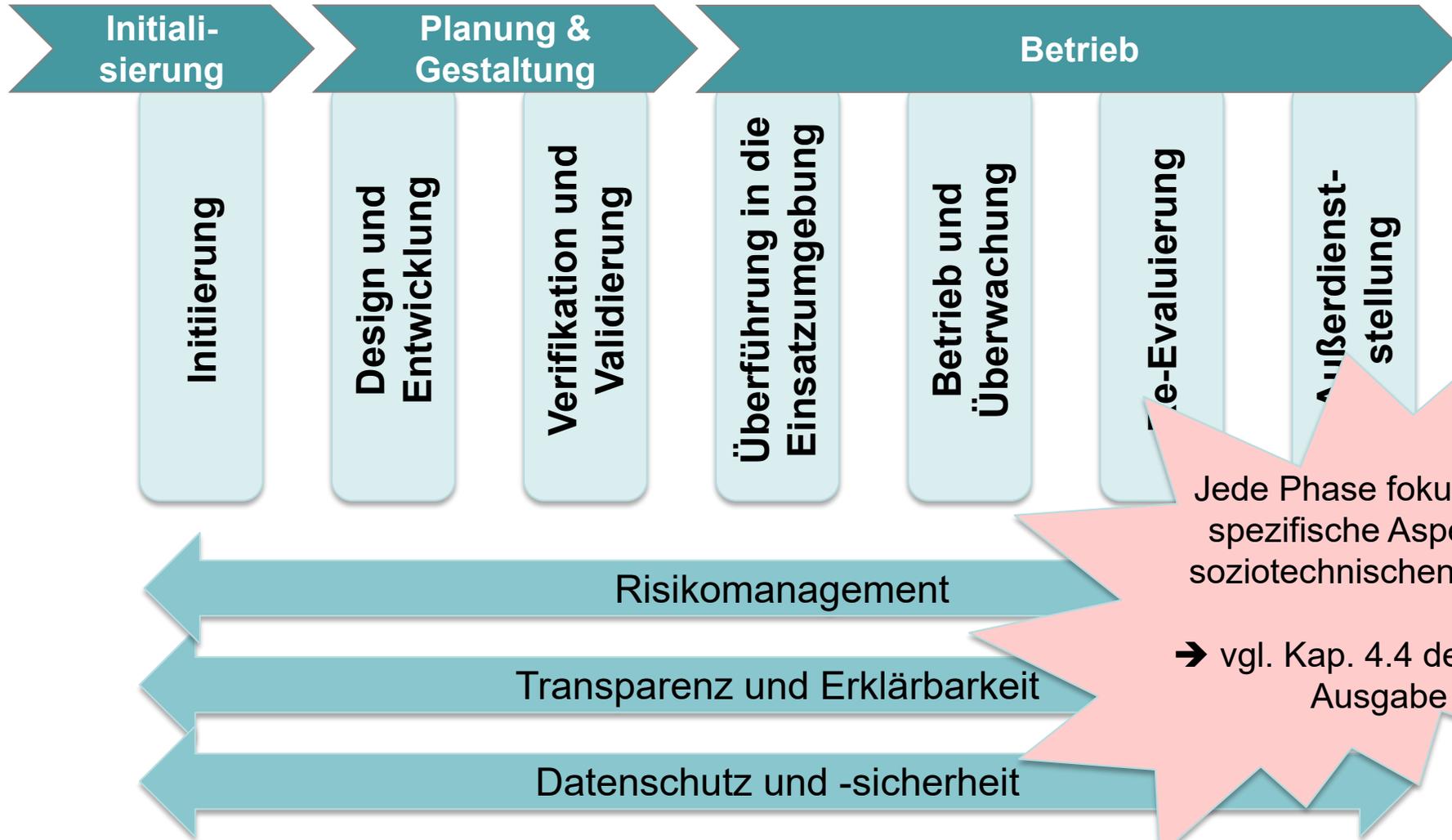
Quelle: Normungsroadmap KI, Ausgabe 2

Ein KI-System ist immer ein sozio-technisches System!

Zur Vertiefung:
Kap. 4.4 der NRM KI,
Ausgabe 2



Kein punktuelle Betrachtung von KI, sondern ganzheitlich über den Lebenszyklus



Jede Phase fokussiert auf spezifische Aspekte des soziotechnischen Systems

→ vgl. Kap. 4.4 der NRM KI, Ausgabe 2

nach ISO/IEC FDIS 22989:2022; Normungsroadmap KI, Ausgabe 2

Den Menschen als Teil des Systems begreifen, und zwar in allen Phasen des KI-Lebenszyklus

Im aktuellen Entwurf der Europäischen Verordnung zu Künstlicher Intelligenz (AI Act) werden insbesondere an Hochrisikosysteme umfangreiche Forderungen zur Einbindung von Menschen gestellt, z. B. Transparenz für Betroffene und Beteiligte, menschliche Aufsicht in unterschiedlichen Rollen und Eingriffsmöglichkeiten bis hin zu einer „Stopptaste“, die von Menschen ausgelöst wird. Welche Transparenz in welchem Kontext für welche Zielgruppe ausreichend ist, wie die menschliche Aufsicht umgesetzt werden sollte und welche Basisinformationen als Grundlage für menschliche Eingriffe ins System vorhanden sein müssen – all das sind Fragestellungen, die vom Menschen aus zu denken und wonach die technischen und sozialen Komponenten zu entwickeln und auszurichten sind.

Um diese soziotechnischen Aspekte in KI-Systemen umzusetzen, sind folgende Herausforderungen zu bewältigen:

- **Angemessenheit:** Technische Komponenten sind auf Grundlage der soziotechnischen Anforderungen auszuwählen.
- **Partizipation:** Die Definition und Auswahl von relevanten Akteur*innen, die beteiligt werden sollten, sind zu operationalisieren.
- **Ethik:** Gesellschaftliche und ethische Fragestellungen mithilfe etablierter Modelle sind zu operationalisieren, messbar bereits bei der Entwicklung der Technologie zu verankern und dabei auf dem Stand der Forschung zu Diskriminierungssensibilität aufzubauen.
- **Kultur:** Eine adäquate Organisationskultur ist zu etablieren (im Arbeitskontext z. B. die Unternehmens-kultur), denn auch diese muss beim KI-Einsatz mitentwickelt werden. Dafür sind relevante Akteur*innen zu sensibilisieren, zu qualifizieren und in einem geeigneten Change Management im Prozess mitzunehmen.
- **Tools:** Über den Lebenszyklus eines KI-Systems hinweg ist der Mensch mit Prozessen, Methoden und Tools zu unterstützen – von der Zielsetzung über die Entwicklung bis zum Betrieb mit Iterationen und Re-Validierung.

[...]

Quelle: Normungsroadmap KI, Ausgabe 2

Das REFA-Institut: Kontakt

Dr.-Ing. Patricia Stock
patricia.stock@refa-institut.de

REFA Fachverband e.V.
REFA-Institut
Emil-Figge-Straße 43
44227 Dortmund

Telefon: +49 231 97 96-213

Fax: +49 231 97 96-219

Internet: www.refa-institut.de

