



Themenkartenstapel: Kriterien einer auf den Menschen ausgerichteten Künstlichen Intelligenz

Themenkarten

Warum muss der Mensch im Mittelpunkt stehen?	3
Kriterium 1: Erklärbarkeit	6
Kriterium 2: Entscheidung durch den Menschen.....	8
Kriterium 3: Mitbestimmung und Akzeptanz	10
Kriterium 4: Technische Robustheit.....	11
Kriterium 5: Datensicherheit und Datenschutz.....	12
Kriterium 6: Verantwortung und Haftung.....	13
Kriterium 7: Diskriminierungsfreiheit	14
Kriterium 8: Ökologische Verantwortung	15
Wie finden die hKI-Kriterien zusammen?	16
Literatur.....	17

Warum muss der Mensch im Mittelpunkt stehen?

Humanzentrierte Künstliche Intelligenz (hKI) ist auf den Menschen und seine Bedürfnisse ausgerichtet. Sie orientiert sich nicht primär an technischer Effizienz, Autonomie und Produktivität, ohne das soziale und ökologische Umfeld ihres Einsatzgebietes auf Augenhöhe zu betrachten. Um die Zentrierung auf den Menschen sicherzustellen, können diverse Kriterien herangeführt werden. Die folgende Einleitung und die Abbildung geben eine kurze Übersicht über mögliche Definitionsansätze und jeweils ein Beispiel, welches das Kriterium in der Praxis verdeutlicht. Die anschließenden Themenkarten gehen näher auf die einzelnen Kriterien ein. Die Kriterien werden ebenfalls für die Sozialpartner der chemischen Industrie mitsamt von Ansätzen des KI-Change-Managements und Handlungsoptionen für die betriebliche Praxis in einem [Handlungsleitfaden](#) aufbereitet.

Infobox – Kriterien einer humanzentrierten KI



Erklärbarkeit und Transparenz



Entscheidung durch den Menschen



Mitbestimmung und Akzeptanz



Technische Robustheit und Risikomanagement



Datensicherheit und Privatsphäre



Verantwortlichkeit, Haftung und Rechtsstaatlichkeit



Diversität, Fairness und Nicht-Diskriminierung



Ökologische und soziale Verantwortung

1. Erklärbarkeit und Transparenz von KI

- **Definition:** Die Prozesse und Entscheidungen der KI sollten für Anwender:innen nachvollziehbar und verständlich sein. Erklärbarkeit bedeutet, dass das System seine Entscheidungen oder Vorhersagen plausibel begründen und darstellen kann.
- **Beispiel:** In einer Software für Logistikanwendungen kommt ein KI-System zu Einsatz, welches die Nachfrage nach einem bestimmten Gut statistisch „voraussagen“ soll. Wenn dem zuständigen Personal transparent dargestellt wird, welche Parameter bei der Berechnung zum Einsatz kommen und wie diese miteinander verknüpft werden, kann der Mensch besser nachvollziehen, wie die KI zu ihrer Handlungsempfehlung kommt.

2. Entscheidung durch den Menschen

- **Definition:** In sensiblen Entscheidungen bleibt der Mensch in der Kontrollposition und kann die KI-Entscheidungen überstimmen oder anpassen. Dies wird als „Human-in-the-Loop“ bezeichnet.
- **Beispiel:** In der medizinischen Diagnostik liefert eine KI Diagnosenvorschläge, jedoch trifft die finale Entscheidung über die Behandlung medizinisches Fachpersonal, welches Empfehlungen der KI prüfen und ablehnen kann, wenn sie nicht passen.

3. Mitbestimmung und Akzeptanz

- **Definition:** Beschäftigte und Stakeholder sollten die Möglichkeit haben, an der Entwicklung und Einführung von KI-Systemen teilzuhaben und ihre Bedenken oder Anforderungen einzubringen, um die Akzeptanz und Nutzen der Technologie zu erhöhen.
- **Beispiel:** Bei der Einführung eines KI-gestützten Zeiterfassungssystems wird der Betriebsrat eingebunden, um sicherzustellen, dass die Bedürfnisse und Erwartungen der Beschäftigten berücksichtigt werden. Das schafft Vertrauen und sorgt für Akzeptanz im Team.

4. Technische Robustheit

- **Definition:** KI-Systeme müssen zuverlässig und fehlerresistent sein, um ihre Funktion unter verschiedenen Bedingungen zu erfüllen. Technische Robustheit reduziert das Risiko, dass Systeme durch unvorhergesehene Eingaben oder Fehler falsch reagieren.
- **Beispiel:** Ein KI-Modell in einem autonom fahrenden Auto muss auch bei schlechten Wetterbedingungen zuverlässig funktionieren. Regelmäßige Tests und Anpassungen sorgen dafür, dass das System Sicherheitsstandards erfüllt.

5. Datensicherheit und Datenschutz

- **Definition:** Die Verarbeitung personenbezogener Daten durch KI-Systeme muss den geltenden Datenschutzbestimmungen entsprechen und Datensicherheit gewährleisten. Nutzer:innen sollen darauf vertrauen können, dass ihre Daten sicher und verantwortungsvoll behandelt werden.
- **Beispiel:** Ein Unternehmen, das KI zur Analyse von Kundenverhalten einsetzt, speichert persönliche Informationen verschlüsselt und anonymisiert Daten, um die Privatsphäre der Kund:innen zu schützen und Sicherheitslücken zu vermeiden.

6. Verantwortung und Haftung

- **Definition:** Es muss klar geregelt sein, wer für die Entscheidungen und Fehler der KI-Systeme verantwortlich ist. Dies betrifft sowohl ethische als auch rechtliche Aspekte.
- **Beispiel:** Wenn eine KI zur Optimierung von Produktionsabläufen eingesetzt wird und eine Reihe an Produkten Fehler aufweisen, darf die Verantwortung nicht auf die Technik abgewälzt werden. Bei der Einführung von KI-Modellen ist zu klären, wer die rechtliche Verantwortung für das Handeln der KI inne hat.

7. Diskriminierungsfreiheit

- **Definition:** KI-Systeme sollten so gestaltet sein, dass sie keine Vorurteile oder Benachteiligungen gegenüber bestimmten sozialen Gruppen aufweisen.
- **Beispiel:** Eine KI, die Stellenbewerbungen vorsortiert, wird so trainiert, dass sie keine Voreingenommenheiten bzgl. Geschlechter-, Alters- oder Herkunftsmerkmalen aufweist. Regelmäßige Überprüfungen und Anpassungen helfen sicherzustellen, dass das System neutral bleibt.

8. Ökologische Verantwortung

- **Definition:** KI-Entwicklung und -anwendung sollten umweltbewusst gestaltet sein und möglichst wenig Ressourcen verbrauchen. Energieeffiziente Modelle und optimierte Datenverarbeitungsprozesse tragen zur ökologischen Verantwortung bei.

Themenkarte 1

- **Beispiel:** Ein Unternehmen verwendet energieeffiziente KI-Modelle und cloudbasierte Server, die mit erneuerbarer Energie betrieben werden, um den ökologischen Fußabdruck zu reduzieren. Auch bei der Datenverarbeitung und dem KI-Training wird darauf geachtet, den Ressourcenverbrauch zu minimieren.

Kriterium 1: Erklärbarkeit

Im Zeitalter der Digitalisierung und der fortschreitenden Entwicklung von Künstlicher Intelligenz gewinnen die Themen **erklärbare KI (XAI)** und menschenzentrierte KI zunehmend an Bedeutung. Insbesondere im wirtschaftlichen und industriellen Sektor, wo KI immer mehr zum Alltag gehört, ist es von zentraler Bedeutung, wie KI-Systeme gestaltet und eingesetzt werden. Dabei ist erklärbare KI nicht nur ein technisches Problem, sondern auch eine Frage der ethischen Verantwortung und der Akzeptanz durch die Menschen, die mit diesen Technologien arbeiten oder von ihnen betroffen sind.

Ein zentraler Aspekt in der Integration von KI in die Arbeitswelt ist die Frage, wie die Menschen mit diesen Technologien interagieren und in welchen Bereichen sie den Maschinen vertrauen. Vertrauen ist ein Schlüsselfaktor, um die Akzeptanz von KI in der Praxis zu gewährleisten. Doch Vertrauen entsteht nicht automatisch – insbesondere dann nicht, wenn die Entscheidungsprozesse der KI intransparent und undurchsichtig sind (KI als sog. „Black Box“).

Die Bedeutung erklärbarer Künstlicher Intelligenz

Hier kommt der Begriff der erklärbaren KI ins Spiel. Erklärbare KI bezieht sich auf Technologien, die es ermöglichen, den Entscheidungsprozess einer KI in einer für Menschen verständlichen Weise darzustellen. Das bedeutet, dass nicht nur das „Was“ einer Entscheidung – etwa das Ergebnis einer Klassifikation oder eine Vorhersage – nachvollziehbar ist, sondern auch das „Warum“ – also die Gründe und Mechanismen, die zu dieser Entscheidung geführt haben. Diese Transparenz ist entscheidend, um das Vertrauen der Nutzer:innen in die Technologie zu gewinnen, insbesondere in einem Arbeitsumfeld, das zunehmend von Automatisierung und KI geprägt wird. Hier helfen zusätzliche **Displays, Anzeigen und die generelle Herangehensweise, Erklärbarkeit bereits in der Gestaltung des KI-Systems** zu bedenken. In der Forschung, wie etwa beim [Fraunhofer Institut IOSB](#), beim [Technologieunternehmen IBM](#) oder in der [Chemiebranche](#) wird KI-Ansätze, die einen besonderen Fokus auf Erklärbarkeit legen, immer mehr Aufmerksamkeit geschenkt. Als allgemeine Differenzierung im Bezug zu XAI wird oftmals zwischen Visualisierungen, schriftlichen Merkmalserläuterung und Erklärverfahren zur Entscheidungsfindung unterschieden.

In der chemischen Industrie werden in Reaktoren Produkte und Vorprodukte für viele weitere Branchen, bspw. für die Automobil- oder pharmazeutischen Industrie, hergestellt. KI-Modelle könnten diese Verfahren überwachen, steuern oder in der Prototypenentwicklung unterstützen. In einem Szenario, in dem eine Störung auftritt, liegt es hier am Grad der Erklärung, wie der Mensch über die Störung benachrichtigt wird. Eine einfache Warnung gibt relativ wenig Informationen preis. Eine KI-Umgebung, die einen besonderen Wert auf Erklärbarkeit der Ergebnisse legt, zeigt umfangreichere Hinweise an. Anstatt lediglich eine Warnung auszugeben, kann eine erklärbare KI detailliert aufzeigen, warum eine Abweichung erkannt wurde, bspw. durch Visualisierungen, die zeigen, wie aktuelle Werte von normalen Mustern abweichen. Die Fachkraft kann so nachvollziehen, auf welcher Grundlage die Entscheidung getroffen wurde, und fundiert entscheiden, ob sie der Empfehlung der KI folgen oder eine andere Maßnahme ergreifen möchten.

Erklärbarkeit ist ein zentraler Bestandteil der menschenzentrierten KI, die den Menschen in den Mittelpunkt stellt und deren Bedürfnisse berücksichtigt. Eine menschenzentrierte KI fördert nicht nur die Effizienz und Leistung von Unternehmen, sondern unterstützt auch die Menschen, die mit dieser Technologie arbeiten, indem sie deren Autonomie, Kontrolle und Verständnis über die Technologien, mit denen sie interagieren, stärkt. Erklärbare KI ist dabei ein entscheidender Baustein, um diese menschenzentrierte Perspektive zu verwirklichen.

Themenkarte 2

Darüber hinaus geht es bei erklärbarer KI nicht nur um die Nachvollziehbarkeit von Entscheidungen, sondern auch um die Möglichkeit, diese zu hinterfragen und gegebenenfalls anzupassen. In einer zunehmend von KI durchdrungenen Arbeitswelt sind Fachkräfte auf ihre Fähigkeit angewiesen, mit den Technologien zu interagieren, sie zu verstehen und gegebenenfalls zu korrigieren. Ohne diese Möglichkeit der Interaktion und Anpassung verlieren KI-Systeme an Akzeptanz und Vertrauen.

Kriterium 2: Entscheidung durch den Menschen

KI verändert die Entscheidungsfindung im Privaten (Navigationssysteme, Suchmaschinen, Übersetzungstools) wie in der Arbeitswelt. In Bereichen der Wirtschaft und Industrie übernimmt KI zunehmend Aufgaben, die früher menschlicher Entscheidungskraft unterlagen. Doch trotz der enormen Effizienz und Präzision moderner KI-Systeme bleibt das Konzept der „Entscheidung durch den Menschen“ ein zentraler Bestandteil einer der Humanzentrierung. Dieses Prinzip betont, dass eine KI transparent aus den [vorhandenen Daten Handlungsempfehlungen](#) abgibt, der Mensch in kritischen Entscheidungsprozessen immer die letzte Instanz bleibt. Besonders im Hinblick auf den Wandel der Arbeitswelt in Unternehmen ist es entscheidend, wie dieses Prinzip mit der Automatisierung und der Unterstützung durch KI zusammengebracht wird.

Der Wandel von Arbeitswelten und die Rolle der KI

Die Arbeitswelt in Wirtschafts- und Industrieunternehmen ist zunehmend von Automatisierung und KI durchzogen. KI wird eingesetzt, um Entscheidungsprozesse zu unterstützen, Routineaufgaben zu automatisieren und das Arbeitsumfeld effizienter zu gestalten. In Bereichen wie der Fertigung, Logistik, dem Kundenservice oder auch der Finanz- und Personalwirtschaft sind KI-Systeme mittlerweile unverzichtbar geworden. Während dies auf der einen Seite enorme Potenziale birgt – etwa durch Zeit- und Kosteneinsparungen – wirft es auf der anderen Seite [wichtige ethische und praktische Fragen](#) auf, insbesondere die, wie viel Entscheidungskompetenz von Maschinen übernommen werden darf und in welchen Fällen der Mensch immer noch die finale Entscheidung treffen muss.

Das Prinzip „Entscheidung durch den Menschen“, als Teil eines humanzentrierten Ansatzes, stellt sicher, dass trotz der fortschreitenden Automatisierung der Mensch die Kontrolle behält, insbesondere bei komplexen oder ethisch sensiblen Entscheidungen. In einem zunehmend automatisierten Arbeitsumfeld ist dies von zentraler Bedeutung, um die Verantwortung und die menschliche Urteilskraft in den Mittelpunkt zu stellen.

Warum die Entscheidung durch den Menschen wichtig ist

Obwohl KI in der Lage ist, riesige Datenmengen zu verarbeiten und Muster zu erkennen, ist sie in der Regel nicht in der Lage, komplexe soziale, moralische oder kontextuelle Nuancen vollständig zu verstehen. Diese Aspekte sind jedoch in vielen beruflichen Entscheidungsprozessen von entscheidender Bedeutung. Der Mensch hat die Fähigkeit, ethische Überlegungen, emotionale Intelligenz und die Berücksichtigung individueller Lebensrealitäten in seine Entscheidungen einzubeziehen – Faktoren, die für Maschinen oft schwer fassbar sind. Gerade in komplexen, dynamischen Umfeldern wie der Arbeitswelt, wo Entscheidungen weitreichende Folgen für das Wohl von Menschen und Organisationen haben können, ist es wichtig, dass der Mensch den endgültigen Entscheidungsprozess übernimmt.

Ein Beispiel zeigt sich in der medizinischen Diagnostik, wo KI bereits als Unterstützung für Ärzt:innen eingesetzt wird, um Diagnosen zu stellen oder Therapieempfehlungen zu geben. Hier hat die KI die Fähigkeit, aus großen Datenmengen zu lernen und Muster zu erkennen, die für den Menschen schwer zugänglich wären. Doch auch wenn die KI eine äußerst präzise Diagnose stellt, bleibt die Ärztin bzw. der Arzt die verantwortliche Person, die letztlich entscheidet, wie die Diagnose in den Kontext der erkrankten Person und ihrer Lebensumstände eingebunden wird. Entscheidungen in der Medizin sind oft nicht nur technischer Natur, sondern auch von ethischen, kulturellen und psychologischen Faktoren geprägt – Aspekte, die der Mensch in seine Entscheidung einfließen lässt.

Themenkarte 3

Das Konzept der „Entscheidung durch den Menschen“ im Kontext menschenzentrierter KI

Im Rahmen menschenzentrierter KI ist das Konzept der „Entscheidung durch den Menschen“ nicht nur ein praktisches Gebot, sondern auch eine ethische Notwendigkeit. Die KI wird nicht als Ersatz für den Menschen betrachtet, sondern als ein Werkzeug, das den Menschen unterstützt und ihm hilft, fundierte Entscheidungen zu treffen. Dabei wird davon ausgegangen, dass der Mensch immer die Verantwortung für die Konsequenzen dieser Entscheidungen trägt.

Eine menschenzentrierte KI fördert die Zusammenarbeit zwischen Mensch und Maschine, wobei die Maschine dem Menschen nicht nur Empfehlungen liefert, sondern ihm auch Entscheidungshilfen und alternative Perspektiven aufzeigt. Der Mensch bleibt der Entscheider, da er die Fähigkeit hat, komplexe, kulturelle und ethische Faktoren in seine Entscheidungen einzubeziehen, die für die KI oft nicht nachvollziehbar sind.

Kriterium 3: Mitbestimmung und Akzeptanz

Besonders in einer menschenzentrierten KI, die den Menschen als Hauptakteur und nicht nur als Nutzer betrachtet, spielt die Einbindung der Beschäftigten in den Entwicklungs- und Implementierungsprozess von KI-Systemen eine wichtige Rolle. Nur wenn Mitarbeiter:innen aktiv in den Einsatz von KI einbezogen werden, steigt die Wahrscheinlichkeit, dass diese Technologien in der Arbeitswelt erfolgreich angenommen werden und tatsächlich einen Mehrwert bieten.

Mitbestimmung bedeutet, dass Beschäftigte die Möglichkeit haben, ihre **Perspektiven und Bedenken bezüglich KI-Anwendungen zu äußern** und mitzugestalten, wie diese in ihrem Arbeitsumfeld eingesetzt werden. Diese Beteiligung fördert nicht nur das Vertrauen in die Technologie, sondern sorgt auch dafür, dass KI-Systeme den tatsächlichen Bedürfnissen der Beschäftigten entsprechen. Ohne **Mitbestimmung und Mitgestaltung** könnte es zu Widerstand oder Ängsten gegenüber der Technologie kommen, insbesondere wenn die KI als Bedrohung für Arbeitsplätze oder die Arbeitsweise wahrgenommen wird. Ein konkretes Vorgehensmodell zur Einbindung der Beschäftigten wird bspw. in einer [aktuellen Interviewstudie](#) vorgeschlagen.

Ein Beispiel aus der industriellen Produktion: In einer Fertigungsanlage wird ein KI-gestütztes System eingeführt, das die Produktionsprozesse überwacht und Optimierungsvorschläge macht. Werden die Mitarbeiter:innen frühzeitig in den Entscheidungsprozess einbezogen, etwa durch Schulungen oder die Möglichkeit, Feedback zu geben, steigt die Akzeptanz des Systems erheblich. Sie fühlen sich nicht nur als passive Nutzer:innen, sondern als aktive Gestalter:innen des Prozesses. In diesem Fall kann die Akzeptanz der neuen Technologie durch das Gefühl der Mitbestimmung entscheidend erhöht werden, was nicht nur die Effektivität des KI-Systems, sondern auch die Arbeitszufriedenheit der Mitarbeiter:innen steigert.

Letztlich ist die Mitbestimmung der Schlüssel, um KI-Anwendungen in der Arbeitswelt nicht nur technologisch, sondern auch sozial erfolgreich zu gestalten. Ein solcher integrativer Ansatz sorgt dafür, dass KI als unterstützendes Werkzeug und nicht als Ersatz wahrgenommen wird, was zu einer höheren Akzeptanz und einem nachhaltigen Wandel der Arbeitswelt führt.

Kriterium 4: Technische Robustheit

Die technische Robustheit von KI-Anwendungen ist daher ein Schlüssel, um den Wandel der Arbeitswelt in Wirtschafts- und Industrieunternehmen erfolgreich zu gestalten. Sie stellt sicher, dass KI-Systeme den Anforderungen der realen Welt gerecht werden und die Menschen, die mit ihnen arbeiten, sich auf die Technologie verlassen können. Nur wenn KI-Systeme stabil und verlässlich sind, kann die Integration in bestehende Arbeitsprozesse ohne größere Risiken oder Störungen erfolgen.

Im Rahmen der menschenzentrierten Künstlichen Intelligenz (KI) wird die technische Robustheit zu einem entscheidenden Faktor für den Erfolg und die Akzeptanz von KI-Systemen in der Arbeitswelt. In der [europäischen KI-Verordnung](#) wird die Widerstandsfähigkeit als Eigenschaft vorgesehen und insbesondere von sog. **hochriskanten KI-Systemen** gefordert. Technische Robustheit bezeichnet die Fähigkeit eines KI-Systems, zuverlässig und stabil zu arbeiten, selbst unter wechselnden Bedingungen, bei Störungen oder in komplexen, unvorhersehbaren Umfeldern. In Unternehmen, in denen KI zunehmend in **kritische Prozesse** integriert wird, ist es von zentraler Bedeutung, dass diese Systeme sowohl leistungsfähig als auch widerstandsfähig gegenüber Fehlern oder unerwarteten Situationen sind. Des Weiteren sind im Idealfall Notfallpläne vorhanden und ungewollte Rückkopplungsschleifen, also das Lernen aus eigens (fehlerhaft) produzierten Daten, wird vermieden. Nur so kann die Technologie das Vertrauen von Mitarbeiter:innen und Führungskräften gewinnen und einen echten Mehrwert schaffen.

Als **Beispiel** für die Bedeutung der technischen Robustheit ist im Bereich der Produktion zu finden. Viele Unternehmen setzen KI-gesteuerte Systeme zur Qualitätskontrolle ein, die automatisch Produkte auf Mängel überprüfen. Wenn das KI-System jedoch bei unvorhergesehenen Änderungen der Produktionsumgebung – etwa durch Verschleiß an Maschinen oder variierende Rohstoffqualitäten – nicht in der Lage ist, präzise und stabile Ergebnisse zu liefern, können Fehler passieren, die die Produktqualität beeinträchtigen. Ein robustes KI-System würde jedoch auch in solchen Variationen zuverlässige Resultate liefern und damit einen reibungslosen Produktionsablauf sicherstellen. Fehlt diese Robustheit, könnte das Vertrauen der Beschäftigten in das System gestört werden, was tendenziell zu einer Ablehnung der Technologie führt und möglicherweise die gesamte Produktionskette beeinträchtigt.

Kriterium 5: Datensicherheit und Datenschutz

Im Kontext der menschenzentrierten Künstlichen Intelligenz (KI) sind Datensicherheit und Datenschutz von zentraler Bedeutung, insbesondere im Zuge des digitalen Wandels. Das [Bundesamt für Sicherheit in der Informationstechnik](#) hat dazu in Kurzform Probleme und Maßnahmen im Bereich der sicheren Künstlichen Intelligenz zusammengetragen. KI-Systeme verarbeiten **große Mengen an sensiblen Daten**, die sowohl private Informationen von Mitarbeiter:innenn als auch geschäftsrelevante Unternehmensdaten umfassen können. Um die Akzeptanz und das Vertrauen in KI-Anwendungen zu gewährleisten, müssen entsprechenden Rahmenbedingungen existieren und Unternehmen sicherstellen, dass diese Daten geschützt sind und die Privatsphäre der betroffenen Personen gewahrt bleibt. Nur durch den konsequenten Schutz der Daten können die potenziellen Vorteile von KI, wie Effizienzsteigerung und Automatisierung, ohne die Gefahr von Missbrauch oder Datenlecks realisiert werden.

Ein **Beispiel** zeigt sich im [Einsatz von KI im Personalmanagement](#) bzw. in der Rekrutierung: Viele Unternehmen setzen KI bereits ein, um Bewerbungen zu analysieren und die besten Kandidaten:innen vorauszuwählen. Solche Praktiken werden kontrovers diskutiert, da sie sowohl Chancen bieten, aber auch vorhandene Diskriminierungsstrukturen reproduzieren können (siehe Abschnitt Diskriminierungsfreiheit). Diese KI-Systeme arbeiten mit personenbezogenen Daten wie Lebensläufen, Qualifikationen und beruflichen Erfahrungen. Wenn jedoch die Datensicherheit nicht gewährleistet ist, bspw. durch geeignete technische und organisatorische Maßnahmen, könnten diese sensiblen Informationen gestohlen oder unrechtmäßig verwendet werden, was zu schwerwiegenden rechtlichen und ethischen Konsequenzen führen würde. Zudem erweckt dies in Beschäftigten und Bewerber:innen das Gefühl, dass ihre Daten ohne ausreichende Kontrolle oder Transparenz verarbeitet werden, was das Vertrauen in das Unternehmen und die Technologie mindert.

Daher sind Datensicherheit und Datenschutz nicht lediglich gesetzliche Anforderungen, sondern auch grundlegende Voraussetzungen für die erfolgreiche Einführung von KI in Unternehmen. Sie sichern die Integrität der verwendeten Daten und schützen sowohl individuelle Rechte als auch die Reputation und das Image des Unternehmens. Der verantwortungsvolle Umgang mit Daten ist ein Baustein der Humanzentrierung, durch den Unternehmen den Wandel der Arbeitswelt mit KI nachhaltig und ethisch vertretbar gestalten können.

Kriterium 6: Verantwortung und Haftung

Im Zuge der Verbreitung von KI in Wirtschaft und Gesellschaft drängen sich zunehmend **rechtliche Fragen** auf. So wird die Klärung von Verantwortungs- und Haftungsfragen zu einem wichtigen Aspekt einer menschenzentrierten KI. KI-Systeme, die in komplexen Arbeitsprozessen eingesetzt werden, können sowohl Chancen als auch Risiken bergen – von der Optimierung von Geschäftsprozessen bis hin zu unerwarteten Fehlern, die zu Schäden führen können. Daher ist es entscheidend, dass Unternehmen klare Regelungen zu Verantwortung und Haftung definieren, um rechtliche und ethische Unsicherheiten zu vermeiden und die Verantwortlichkeit in der Nutzung von KI zu gewährleisten. In der EU wurde im Jahr 2022 eine KI-Haftungsrichtlinie verabschiedet, die Angaben für zivilrechtliche Haftungsregelungen gibt. Eine [aktuelle Studie \(Stand November 2024\) des Wissenschaftlichen Diensts des Europäischen Parlaments](#) kommt zu dem Schluss, dass die Richtlinie auf alle Softwareprodukte ausgeweitet werden sollte.

Ein **Beispiel** verdeutlicht die Bedeutung dieser Klärung: In einem Fertigungsunternehmen wird ein KI-gestütztes System zur Überwachung und Steuerung von Produktionsmaschinen eingesetzt. Angenommen, das System trifft eine falsche Entscheidung, die zu einem Maschinenfehler führt, der in der Folge einen Produktionsstop und hohe Kosten verursacht. In einem solchen Fall muss klar definiert werden, wer für den Fehler haftet – ist es das Entwicklungsteam des KI-Systems, die das System nicht ausreichend getestet haben, oder Führungskräfte der Betreiberfirma, die das System möglicherweise falsch einsetzen? Wenn solche Haftungsfragen im Vorfeld nicht geklärt sind, können langwierige rechtliche Auseinandersetzungen und Unsicherheiten darüber entstehen, wer für den Schaden verantwortlich ist.

Die **Klärung von Verantwortungs- und Haftungsfragen** ist daher essenziell, um Vertrauen in die Technologie zu schaffen und die Integration von KI in Unternehmen zu erleichtern. Sie sorgt dafür, dass alle Beteiligten – von Entwickler:innen bis zu Nutzer:innen – ihre Rollen und Pflichten verstehen und mögliche rechtliche Risiken minimiert werden. Auf diese Weise kann die Arbeit mit KI-Systemen sicherer und transparenter gestaltet werden, was für den Wandel der Arbeitswelt in Unternehmen von zentraler Bedeutung ist.

Kriterium 7: Diskriminierungsfreiheit

Um sicherzustellen, dass KI-Systeme **fair**, **transparent** und **inklusiv** gestaltet werden, ist das Prinzip der Nicht-Diskriminierung im Rahmen der menschenzentrierten Künstlichen Intelligenz von entscheidender Bedeutung. KI-Anwendungen in der Industrie haben, wie bereits dargestellt, das Potenzial, Entscheidungsprozesse zu automatisieren, etwa bei der Personalgewinnung und -planung, der Prozessoptimierung oder der Kundenbetreuung. Werden diese Systeme jedoch nicht sorgfältig entwickelt, können sie unbewusste Vorurteile und Stereotype (bekannt als Bias) verstärken und unfaire Diskriminierung nach Kriterien wie Geschlecht, Alter, ethnischer Herkunft oder Behinderung zur Folge haben. Weitere Informationen, Hinweise und Erläuterungen finden sich bspw. auf der Seite der [Hans-Böckler-Stiftung](#) oder in einer [Veröffentlichung der Plattform Lernende Systeme](#).

Um dies zu verdeutlichen, kann erneut auf das **Beispiel der Personalrekrutierung** zurückgegriffen werden: Ein Unternehmen nutzt ein KI-gestütztes System zur Auswahl von Bewerbungen. Wenn das KI-System auf historischen Daten basiert, die unbewusste Vorurteile enthalten – etwa eine höhere Zahl an Männern in technischen Positionen oder eine Präferenz für bestimmte Ausbildungswege – kann es dazu führen, dass qualifizierte Bewerber:innen aus unterrepräsentierten Gruppen benachteiligt werden. Das Ergebnis ist eine verzerrte, diskriminierende Auswahl, die nicht nur die Vielfalt im Unternehmen verringert, sondern auch rechtliche und ethische Probleme aufwirft.

Nicht-Diskriminierung ist daher ein zentraler Bestandteil einer menschenzentrierten KI, die sicherstellen soll, dass alle Menschen unabhängig von Geschlecht, Herkunft oder anderen persönlichen Merkmalen gleiche Chancen und Rechte in der Arbeitswelt erhalten. Die Schaffung diskriminierungsfreier KI-Anwendungen fördert nicht nur die Vielfalt und Inklusion, sondern stärkt auch das Vertrauen der Mitarbeiter und der Gesellschaft in die Technologie. In einer zunehmend digitalisierten Arbeitswelt ist es daher entscheidend, dass KI-Systeme fair gestaltet werden, um einen gerechten und transparenten Wandel der Arbeitswelt zu ermöglichen.

Kriterium 8: Ökologische Verantwortung

KI-Systeme bringen nicht nur eine **soziale, sonder auch eine ökologische Verantwortung** mit sich. KI trägt zum Wirtschaftswachstum bei und verbraucht viel Energie. Dies könnte jedoch kompensiert werden durch Beiträge, die KI zum Erreichen der Klimaschutzziele und dem Unterstützen nachhaltiger Produktions- und Lebensweisen leisten kann. KI-Systeme bieten nicht nur die Möglichkeit, Prozesse zu optimieren und Effizienz zu steigern, sondern sie können auch einen erheblichen Einfluss auf die Umwelt haben, sei es durch den Energieverbrauch von Rechenzentren oder durch die Nutzung von Ressourcen für die Entwicklung und den Betrieb dieser Systeme. Eine menschenzentrierte KI berücksichtigt daher nicht nur die sozialen und ethischen Aspekte, sondern auch die ökologischen Auswirkungen, um eine nachhaltige Zukunft zu fördern.

Die Berücksichtigung von **ökologischer Verantwortung im Design von KI-Anwendungen** bedeutet daher nicht nur, auf den Energieverbrauch zu achten, sondern auch die langfristigen Auswirkungen auf die Umwelt zu bedenken und Technologien zu entwickeln, die die Sustainable Development Goals (SDGs) der Vereinten Nationen unterstützen. KI-Systeme, die auf nachhaltige Praktiken ausgerichtet sind – etwa durch energieeffiziente Algorithmen oder die Nutzung erneuerbarer Energien in Rechenzentren – tragen aktiv zu einer Reduktion des ökologischen Fußabdrucks bei und fördern eine ressourcenschonende Wirtschaft. Daneben werden Anwendungen Künstlicher Intelligenz vielfach dazu genutzt, ressourcenschonende Produkte zu optimieren, Vorschläge für effizientere Produktionsprozesse zu erarbeiten oder Extremwettereignisse sowie Waldbrände vorausberechnen. Offenbar geförderten werden etwa sog. Leitprojekte in dem Programm „KI-Leuttürme für Umwelt, Klima, Natur und Ressourcen“ vom Bundesumweltministerium.

Indem Unternehmen ökologische Verantwortung in die Gestaltung ihrer KI-Lösungen integrieren und die Forschung solche Lösungen vorantreibt, können sie nicht nur ihre eigenen Nachhaltigkeitsziele erreichen, sondern auch einen positiven Beitrag zu globalen Umweltzielen leisten. So wird KI zu einem Instrument, das nicht nur den Wandel von Arbeits- und Lebenswelten effizient gestaltet, sondern gleichzeitig den Weg für eine nachhaltige und zukunftsfähige Wirtschaft ebnnet.

Wie finden die hKI-Kriterien zusammen?

Die Integration von KI in Wirtschaft und Gesellschaft verändert grundlegend die Art und Weise, wie Arbeit organisiert wird. KI-Anwendungen optimieren Prozesse, automatisieren Routineaufgaben und ermöglichen es Unternehmen, schneller und effizienter auf Marktveränderungen zu reagieren. Besonders in Bereichen wie Fertigung, Logistik, Kundenservice und Personalmanagement ist der Einfluss von KI unübersehbar. Jedoch geht dieser technologische Wandel nicht ohne Herausforderungen einher, vor allem im Hinblick auf die Art und Weise, wie Arbeitsplätze gestaltet und welche Kompetenzen von den Beschäftigten verlangt werden.

Die **Kriterien einer menschenzentrierten Künstlichen Intelligenz** – wie Erklärbarkeit, Entscheidung durch den Menschen und Mitbestimmung – müssen in der Arbeitswelt zusammen gedacht werden, um einen ganzheitlichen Fokus auf die Bedürfnisse der Menschen und Beschäftigten zu gewährleisten. Jedes dieser Kriterien adressiert einen spezifischen Aspekt der Interaktion zwischen Mensch und KI, doch nur in ihrer Kombination können sie ein Arbeitsumfeld schaffen, das sowohl technologisch fortschrittlich als auch sozial und ethisch verträglich ist.

Erklärbare KI und die Möglichkeit der menschlichen Entscheidungsfindung stellen sicher, dass Beschäftigte KI-Systeme verstehen und deren Entscheidungen hinterfragen können. Mitbestimmung fördert Vertrauen, Akzeptanz und eine aktive Einbindung sowie Mitgestaltung. Technische Robustheit und Datenschutz sorgen dafür, dass KI-Anwendungen sicher und zuverlässig sind und sowohl rechtliche als auch ethische Anforderungen erfüllen. Nicht-Diskriminierung gewährleistet Fairness und Chancengleichheit, während ökologische Verantwortung die Energieeffizienz und langfristige Zukunftsfähigkeit von Unternehmen unterstützt.

Auch wenn die dargestellten Kriterien der Humanzentrierung als Idealbild gelten und in der Realität oftmals nicht alle gleichsam verfolgt werden können, ist ihre Bedeutsamkeit als Orientierungspunkte hervorzuheben. Indem diese Aspekte gemeinsam in der Planung von KI-Systemen berücksichtigt werden, entsteht ein **integrativer Ansatz**, der technologische Innovation und die Bedürfnisse der Menschen miteinander in Einklang bringt. Dies ist essenziell, um die Arbeitswelt so zu gestalten, dass sie nicht nur effizienter und produktiver wird, sondern auch die Würde, Rechte und das Wohlbefinden der Beschäftigten in den Mittelpunkt stellt. Ein solch ganzheitlicher Fokus ist der Schlüssel, um die Potenziale von KI voll auszuschöpfen und gleichzeitig eine gerechte, sichere und nachhaltige Arbeitswelt zu schaffen.



Literatur

Beck, Susanne/Grunwald, Armin/Jacob, Kai/Matzner, Tobias (2019): Künstliche Intelligenz und Diskriminierung – Whitepaper aus der Plattform Lernende System. https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG3_Whitepaper_250619.pdf (Abruf am 21.11.2024).

Beesch, Simon/Malanowski, Norbert/Nisser, Annerose (2023): Mit Künstlicher Intelligenz die Sustainable Development Goals erreichen. Perspektiven für Energiewirtschaft und Luftfahrt, HBS-Forschungsförderung Working Paper Nr. 283. https://www.boeckler.de/fpdf/HBS-008590/p_fofoe_WP_283_2023.pdf (Abruf am 21.11.2024).

Bundesamt für Sicherheit in der Informationstechnik (2021): Sicherer, robuster und nachvollziehbarer Einsatz von KI. Probleme, Maßnahmen und Handlungsbedarfe. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Herausforderungen_und_Massnahmen_KI.pdf?__blob=publicationFile&v=6 (Abruf am 21.11.2024).

Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz (o. D.): Unsere Förderinitiative "KI-Leuchttürme für Umwelt, Klima, Natur und Ressourcen" <https://www.bmuv.de/themen/digitalisierung/ueberblick-digitalisierung/unsere-foerderinitiative-ki-leuchttuerme-fuer-umwelt-klima-natur-und-ressourcen#c40106> (Abruf am 21.11.2024).

Carstensen, Tanja/Ganz, Kathrin (2023): Gender, Künstliche Intelligenz und die Arbeit der Zukunft. Eine Analyse der Aushandlungsprozesse in wissenschaftlichen, medialen und politischen Diskursen und der Möglichkeiten (betrieblicher) Gestaltung, HBS-Forschungsförderung Working Paper Nr. 274. https://www.boeckler.de/data/impuls_2023_04_S3.pdf (Abruf am 21.11.2024).

Chemie.de (2022): Wie erklärbare künstliche Intelligenz das Wachstum der Industrie 4.0 vorantreiben kann. <https://www.chemie.de/news/1177625/wie-erklaerbare-kuenstliche-intelligenz-das-wachstum-der-industrie-4-0-vorantreiben-kann.html> (Abruf am 21.11.2024).

Europäische Kommission (2024): Der AI Act Explorer. Artikel 15: Genauigkeit, Robustheit und Cybersicherheit. <https://artificialintelligenceact.eu/de/article/15/> (Abruf am 21.11.2024).

Fleck, Lara/Rounding, Nicholas/Özgül, Pelin (2022): Künstliche Intelligenz in der Personalauswahl. <https://www.denkfabrik-bmas.de/fileadmin/Downloads/Publikationen/Kuenstliche-Intelligenz-in-der-Personalauswahl.pdf> (Abruf am 21.11.2024).

Fraunhofer Institut für Optronik, Systemtechnik und Bildauswertung IOSB (o. D.): Erklärbare KI. https://www.iosb.fraunhofer.de/de/kompetenzen/bildauswertung/human-ai-interaction/erklaerbare-ki.html#copy_1434514291 (Abruf am 21.11.2024).

Heilig, Thorsten (2024): Entscheidungsintelligent werden mit KI. <https://www.springerprofessional.de/unternehmensprozesse/kuenstliche-intelligenz/entscheidungsintelligent-werden-mit-ki/26789262> (Abruf am 21.11.2024).

IBM (o. D.): Was ist erklärbare KI? <https://www.ibm.com/de-de/topics/explainable-ai> (Abruf am 21.11.2024).

Kompetenzzentrum Öffentliche IT (o. D.): Welche Aufgabe wollen wir Maschinen überlassen? <https://www.oeffentliche-it.de/-/automatisierte-entscheidungen> (Abruf am 21.11.2024).

Malanowski, Norbert/Nisser, Annerose/Auer, Michael/Beesch, Simon (2022): Handlungsleitfaden für die Sozialpartner. Humanzentrierte Künstliche Intelligenz in der chemischen Industrie. https://www.vditz.de/fileadmin/Publikationen/2022-10-18_Handlungsleitfaden_f._e._humanzentrierte_KI.pdf (Abruf am 21.11.2024).

Ruess, Patrick/Staffa, Anna/Kreutz, Anna/Busch, Christine/Saba Gayoso, Christian Oswaldo/Pollmann, Kathrin (2024): Künstliche Intelligenz in betrieblichen Prozessen: Ein Vorgehensmodell zur partizipativen Gestaltung von KI-Anwendungen, in HMD Praxis der



Wirtschaftsinformatik 2/2024. <https://www.springerprofessional.de/kuenstliche-intelligenz-in-betrieblichen-prozessen-ein-vorgehens/26789064> (Abruf am 21.11.2024).

Wold, Jacob Wulff (2024): Studie des EU-Parlaments: KI-Haftungsrichtlinie sollte allgemeine KI abdecken. <https://www.euractiv.de/section/innovation/news/studie-des-eu-parlaments-ki-haftungsrichtlinie-sollte-allgemeine-ki-abdecken/> (Abruf am 21.11.2024).