



# Kriterien für die Mensch-Maschine- Interaktion bei KI

Ansätze für die menschengerechte Gestaltung in der Arbeitswelt

**WHITEPAPER**

Norbert Huchler et al.  
AG Arbeit/Qualifikation,  
Mensch-Maschine-Interaktion

# Inhalt

---

Zusammenfassung .....	3
1. Arbeitsteilung zwischen Mensch und Maschine .....	5
2. Gestaltung der Mensch-Maschine-Interaktion .....	8
2.1 Cluster 1: Schutz des Einzelnen .....	9
2.2 Cluster 2: Vertrauenswürdigkeit .....	13
2.3 Cluster 3: Sinnvolle Arbeitsteilung .....	16
2.4 Cluster 4: Förderliche Arbeitsbedingungen .....	19
3. Umsetzung der Kriterien und Ausblick .....	23
Über dieses Whitepaper .....	27
Literatur .....	29

## Zusammenfassung

Der Einsatz von Künstlicher Intelligenz (KI) eröffnet vielfältige Potentiale für ein sicheres, eigenverantwortliches und selbstbestimmtes Arbeiten sowie attraktive und wettbewerbsfähige Arbeitsplätze. So können KI-basierte Assistenzsysteme Beschäftigte beispielsweise von anstrengenden oder gefährlichen Tätigkeiten entlasten und bei komplexen Prozessen und Entscheidungen unterstützen. Gleichzeitig verändern KI-Systeme die Interaktion von Mensch und Technik in der Arbeitswelt. Mensch und Maschine werden künftig noch stärker – und anders – als bisher zusammenwirken, da Maschinelles Lernen (ML) und ähnliche Technologien Maschinen befähigen, bestimmte Aufgaben selbstständig auszuführen und dabei laufend dazuzulernen.

Die zunehmende Kollaboration zwischen Mensch und Technik macht beim Einsatz von Künstlicher Intelligenz eine Neujustierung der Aufgabenverteilung notwendig. Um diese im Sinne des Menschen zu gestalten, muss die Technologie an die Vorteile und Potentiale menschlichen Denkens und Handelns anknüpfen und die wechselseitige Ergänzung – und nicht Ersatz oder Konflikt – in den Mittelpunkt der Interaktion stellen. Eine koordinierte Balance, die sowohl den Beschäftigten als auch den technologischen und wirtschaftlichen Potentialen der Künstlichen Intelligenz gerecht wird, erhöht die Chancen für einen individuell und gesellschaftlich akzeptierten Einsatz von Künstlicher Intelligenz in der Arbeitswelt.

Nur durch klar definierte Orientierungsmaßstäbe für die neue Aufgabenverteilung können sichere Arbeitsplätze geschaffen, qualifizierte Arbeitskräfte ausgebildet sowie die Gestaltung guter und menschengerechter Arbeit umgesetzt werden. An dieser Stelle setzt das Whitepaper an und legt einen Kriterienkatalog für die Mensch-Maschine-Interaktion im Arbeitskontext vor. Die Kriterien, die auf eine nachhaltig zukunftsorientierte und menschenzentrierte Gestaltung der Mensch-Maschine-Interaktion zielen, lassen sich zu vier Clustern zusammenfassen:

<p><b>Cluster 1: Schutz des Einzelnen</b></p> <ul style="list-style-type: none"> <li>■ Sicherheit und Gesundheitsschutz</li> <li>■ Datenschutz und verantwortungsbewusste Leistungserfassung</li> <li>■ Vielfaltssensibilität und Diskriminierungsfreiheit</li> </ul>	<p><b>Cluster 2: Vertrauenswürdigkeit</b></p> <ul style="list-style-type: none"> <li>■ Qualität der verfügbaren Daten</li> <li>■ Transparenz, Erklärbarkeit und Widerspruchsfreiheit</li> <li>■ Verantwortung, Haftung und Systemvertrauen</li> </ul>
<p><b>Cluster 3: Sinnvolle Arbeitsteilung</b></p> <ul style="list-style-type: none"> <li>■ Angemessenheit, Entlastung und Unterstützung</li> <li>■ Handlungsträgerschaft und Situationskontrolle</li> <li>■ Adaptivität, Fehlertoleranz und Individualisierbarkeit</li> </ul>	<p><b>Cluster 4: Förderliche Arbeitsbedingungen</b></p> <ul style="list-style-type: none"> <li>■ Handlungsräume und reichhaltige Arbeit</li> <li>■ Lern- und Erfahrungsförderlichkeit</li> <li>■ Kommunikation, Kooperation und soziale Einbindung</li> </ul>

Diese Kriterien richten sich an Akteure aus der Planung und Entwicklung Lernender Systeme sowie an Akteure, die an der Implementierung von KI-Systemen in Unternehmen beteiligt sind. Der Kriterienkatalog soll Handreichungen für die Gestaltung der Aufgabenteilung zwischen Mensch und Technik bei der Anwendung Lernender Systeme anbieten. Außerdem sollen die Kriterien zur Weiterentwicklung bestehender Regelungen inspirieren – etwa in der Normung, Gesetzgebung oder Sozialpartnerschaft – und künftig ein flexibleres, selbstbestimmteres und eigenverantwortlicheres Arbeiten ermöglichen.

# 1. Arbeitsteilung zwischen Mensch und Maschine

---

Der Einsatz von Künstlicher Intelligenz (KI) in der Arbeit macht eine Neujustierung der „Arbeitsteilung“ zwischen Mensch und Technik notwendig. Gründe dafür sind die Neuverteilung von Tätigkeiten in Arbeitssystemen allgemein, aber vor allem auch das stärkere Zusammenrücken und die zunehmende direkte Kollaboration an der Schnittstelle zwischen Mensch und Maschine durch Lernende Systeme: Maschinelles Lernen und ähnliche Technologien verstärken die Fähigkeit komplexer technisch-organisationaler Systeme, den Menschen als eine Art von „Akteur“<sup>1</sup> gegenüberzutreten und in der Interaktion die Handlungsträgerschaft für bestimmte Aufgaben übernehmen zu können.<sup>2</sup>

Für diese neue Aufgaben- und Rollenteilung sind konkrete Kriterien erforderlich, die Ansatzpunkte für eine nachhaltig zukunftsorientierte und menschenzentrierte Gestaltung der Mensch-Maschine-Interaktion in der Arbeitswelt formulieren. Wichtig ist dabei ein systematischer Blick auf die unterschiedlichen Potentiale von Mensch und Technik einerseits sowie auf die verschiedenen Interessenlagen der beteiligten Akteure andererseits.

Eine derartige „Arbeitsteilung“ ist umso erfolgreicher, je besser sich die jeweiligen Stärken und Potentiale von Mensch und Technik – also die spezifischen menschlichen und technischen Fähigkeiten und Eigenschaften – in eine sich wechselseitig bestärkende, „co-evolutionäre“ Beziehung setzen lassen (Huchler 2016). Damit die Menschen durch Künstliche Intelligenz in ihrer zentralen Rolle in der Arbeitswelt nicht geschwächt, sondern bestätigt und bestärkt werden, braucht es eine Sensibilität für die Möglichkeiten und Notwendigkeiten einer komplementären Ergänzung zwischen menschlicher Arbeit und teilautonomen intelligenten beziehungsweise Lernenden Systemen.

Eine realistische Auseinandersetzung mit den spezifischen Vorteilen, aber auch den immanenten, in der Technologie und ihrer Anwendung selbst liegenden Grenzen und Defiziten – wie im Falle Künstlicher Intelligenz beispielsweise die Datenabhängigkeit, der fehlende „Sinn“ für beziehungsweise Mangel an Repräsentanz von Kausalitäten, die Gefahr von Zirkelschlüssen oder Pfadabhängigkeiten – sind bei der Entwicklung von Gestaltungskriterien von Bedeutung.

---

1 Da in lernende KI-Systeme immer umfangreichere Interaktionsfähigkeiten „eingeschrieben“ beziehungsweise „einprogrammiert“ und komplexere Aufgabenteilungen an der Schnittstelle möglich werden, wird dem „sozialen Akteur“ ein „technischer Akteur“ gegenübergestellt. Zwischen diesen Akteursformen ist jedoch streng zu unterscheiden – etwa hinsichtlich der Kompetenzen wie auch der rechtlichen Implikationen und dem Tragen von Verantwortung.

2 „Interaktion“ umfasst im Folgenden nicht nur die „soziale Interaktion“ zwischen Menschen, sondern auch die „soziale Interaktion mit Gegenständen“ einerseits und die „in Objekte eingeschriebenen“ beziehungsweise „objekt-vermittelten Interaktionen“ andererseits. Auch wird „Handlungsträgerschaft“ bestimmt als Träger für formal definierte Aufgaben/Funktionen, was sowohl Personen, Organisationen/Institutionen (kollektive Akteure) wie auch technische Objekte sein können.

Für die menschengerecht gestaltete Mensch-Maschine-Interaktion bei Künstlicher Intelligenz ist wichtig, dass die Technologie an menschliches Handeln anschlussfähig ist, und zwar genau dort, wo die besonderen Vorteile und Potentiale menschlichen Denkens und Handelns liegen – wie zum Beispiel das Handeln unter Unsicherheit oder bei unvollständigen Informationen und Widersprüchen, die Kombination von Fachwissen, Erfahrung und implizitem Wissen zu Kompetenzen und Erfahrungswissen oder auch die Möglichkeit, Informationen in Abhängigkeit vom jeweiligen sozialen Kontext Sinn zuzuschreiben („Indexikalität“). Dies stellt hohe Anforderungen an die interaktive Anschlussfähigkeit beziehungsweise „komplementäre Adaptivität“ (Huchler 2019) des technischen Systems – um die wechselseitige Ergänzung und nicht den Ersatz beziehungsweise Konflikt in den Mittelpunkt der Schnittstellengestaltung zu stellen.

Die Potentiale und Anforderungen einer menschenzentrierten Technikentwicklung und Arbeitsgestaltung müssen mit den Anforderungen und Möglichkeiten der ökonomischen Rationalisierung und des technologischen Fortschritts verhandelt und abgestimmt werden. Hintergrund und Ausgangspunkt sind dabei die geltenden rechtlichen Bestimmungen – wie zum Beispiel Arbeitsrecht und Arbeitsschutzgesetze, Schutz der Persönlichkeitsrechte und Datenschutzbestimmungen oder Regelungen zur Mitbestimmung. Eine koordinierte Balance, die sowohl den Beschäftigten und ihrer Entwicklung als auch den Potentialen technologischer Innovationen für den Menschen und die Gesellschaft gerecht wird, ermöglicht einen angemesseneren und effektiveren Einsatz von Künstlicher Intelligenz im betrieblichen Umfeld (inklusive der gesellschaftlichen Folgen) sowie am Standort Deutschland allgemein.

Eine Gestaltungsperspektive, die ausgehend von den unterschiedlichen Kompetenzen zu einer optimalen „Arbeitsteilung“ an der Schnittstelle zwischen Mensch und technischem System kommt, bietet einen realistischen Blick auf Einsatzszenarien, fördert die Praxistauglichkeit und Zukunftsfähigkeit von Entwicklungen und könnte nicht zuletzt dazu beitragen, Vorbehalte und Bedenken abzubauen sowie soziale wie technische Innovationen voranzubringen.

Die Entwicklung von Kriterien für die Ausgestaltung der Mensch-Maschine-Interaktion bei Lernenden Systemen ist ein wichtiges Element im Hinblick auf die Einführung dieser Systeme in der Arbeitswelt. Geeignete Gestaltungskonzepte sind ein wesentlicher Baustein für die Umsetzung der Transformationsprozesse in den Betrieben. Schon bei der Entwicklung technischer Systeme – und später beim Einsatz Lernender Systeme – müssen für die Gestaltung der Mensch-Maschine-Interaktion passende „arbeitsteilige“ Einsatzszenarien reflektiert werden.

Mit Blick auf die Akzeptanz der Beschäftigten und die Passung auf den konkreten Arbeitsprozess ist zudem ein partizipatives Vorgehen sowie ein sozialpartnerschaftlicher Ansatz vonnöten. Dies bietet einen gemeinsam getragenen, innovationsförderlichen Rahmen für die Ausgestaltung und Einführung von KI-Technologien und Lernenden Systemen.

Die Entwicklung von Kriterien für die Gestaltung der Mensch-Maschine-Interaktion im Arbeitskontext steht im Fokus dieses Whitepapers der Arbeitsgruppe 2 – Arbeit/Qualifikation, Mensch-Maschine-Interaktion der Plattform Lernende Systeme. Das Papier richtet sich dabei insbesondere an Akteure aus der Planung und Entwicklung Lernender Systeme sowie an Akteure, die an der Implementierung von KI-Systemen beteiligt sind. Ziel ist es, dass der nachfolgende Kriterienkatalog – zum Beispiel als Reflexionsinstrument – frühzeitig bei technikgetriebenen Innovationsprozessen rund um Künstliche Intelligenz entsprechende Berücksichtigung findet.<sup>3</sup>

---

3 Dabei lehnt sich das künftige Gestaltungsinstrument an andere Modelle an, die als Ansatzpunkte für die Entwicklung von Kriterien für die Gestaltung der Mensch-Maschine-Interaktion herangezogen werden können – exemplarisch: MEESTAR-Modell (**M**odell zur **e**thischen **E**valuation **s**ozio-**t**echnischer **A**rrangements) (Manzeschke et al. 2013).

## 2. Gestaltung der Mensch-Maschine-Interaktion

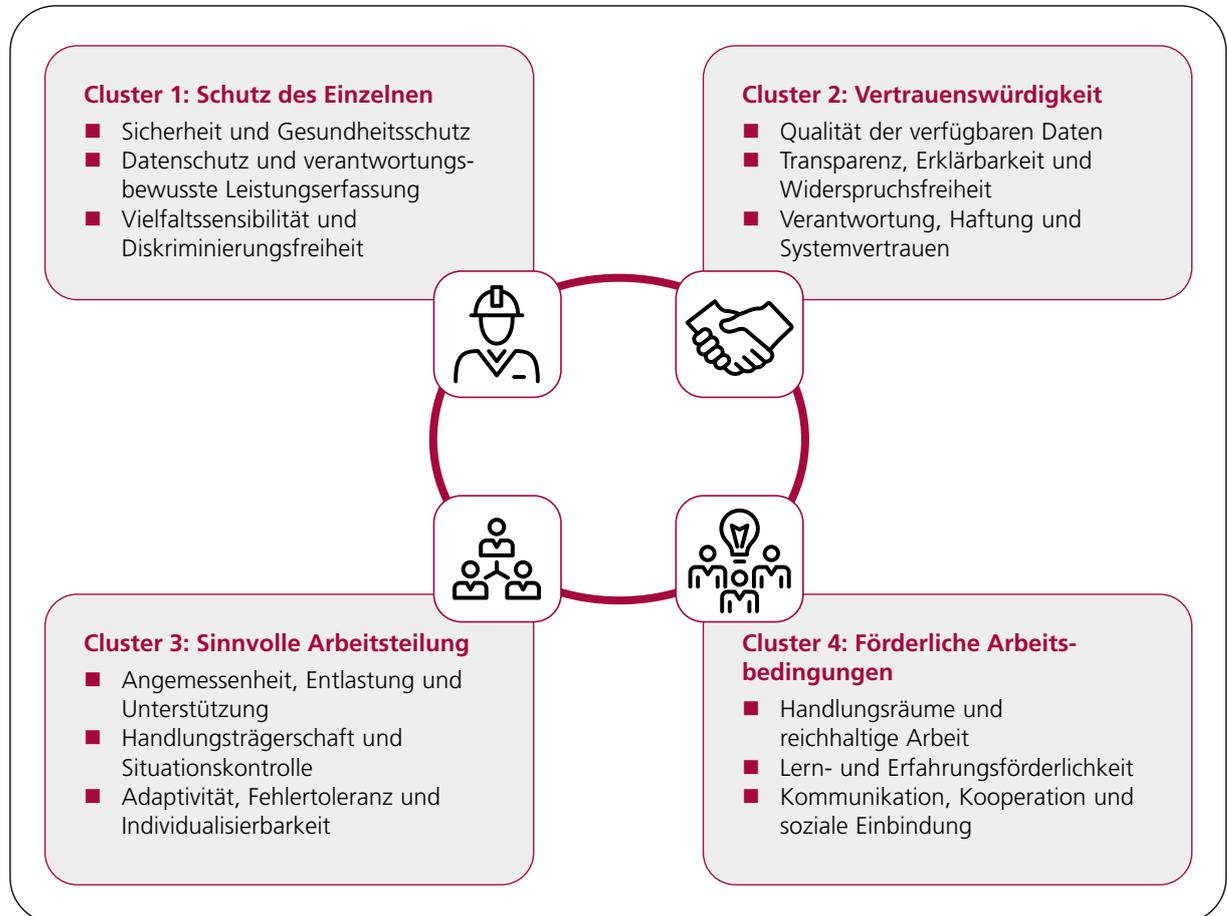
---

Ziel des Kriterienkatalogs ist es, Ansatzpunkte speziell für die Gestaltung der Mensch-Maschine-Interaktion bei Künstlicher Intelligenz zu entwickeln und über bestehende Ansätze für die Gestaltung digitaler Technologien hinauszugehen. Ein wesentliches Unterscheidungsmerkmal besteht dabei in der veränderten „Arbeitsteilung“ und Interaktion an der Schnittstelle zwischen Mensch und KI-basierten Assistenzsystemen, die eine oder mehrfache Übergaben von Handlungsträgerschaft oder sogar ihre Überlagerung zwischen Mensch und Maschine beinhaltet.

Der Kriterienkatalog soll Handreichungen für die technische und betriebliche Gestaltung der „Arbeitsteilung“ zwischen Mensch und Technik bei Künstlicher Intelligenz formulieren sowie die Weiterentwicklung bestehender Regelungen inspirieren – etwa in der Normung, Gesetzgebung oder Sozialpartnerschaft. Ausgangspunkt und Rahmen sind bestehende rechtliche und sozialpartnerschaftliche Regelungen – wie etwa arbeits-, arbeitsschutz-, persönlichkeits- und datenschutzrechtliche Bestimmungen oder gesetzliche Regelungen zur Mitbestimmung. Gleichzeitig sollen die Kriterien auch die Ziele der Produktivität und der Innovationsfähigkeit des Standortes unterstützen sowie flexibleres, reichhaltiges und selbstbestimmteres Arbeiten ermöglichen.

Im Folgenden wird der Kriterienkatalog zunächst im Überblick präsentiert (Abbildung 1). Anschließend werden die einzelnen Kriterien im Detail vorgestellt. Die Kriterien lassen sich in vier Cluster zusammenfassen, die von den Anforderungen an den Schutz des Einzelnen etwa im Hinblick auf Sicherheit und Privacy (Cluster 1) und die Vertrauenswürdigkeit (Cluster 2) bis hin zu Anforderungen an die betriebliche Gestaltung in Bezug auf eine sinnvolle „Arbeitsteilung“ (Cluster 3) und förderliche Arbeitsbedingungen (Cluster 4) reichen. Zu beachten ist dabei, dass die Kriterien vielfältige Querbezüge aufweisen, sodass eine trennscharfe Differenzierung der Kriterien nicht in allen Fällen möglich ist. Zudem ist wichtig, bei der konkreten Entwicklung und beim Einsatz von KI-Systemen in den Unternehmen die unterschiedlichen Automatisierungs- beziehungsweise Autonomielevel sowie Kritikalitätslevel des jeweiligen Anwendungsfalls zu berücksichtigen. Die Kriterien müssen jeweils auf die verschiedenen Niveaus bezogen und mit diesen wechselseitig abgestimmt werden (ähnlich etwa zu anderen Bereichen wie Mobilität oder Industrie: acatech 2016, acatech/Fachforum Autonome Systeme 2017, Plattform Industrie 4.0 2019b). Im Fokus stehen interaktive Mensch-Maschine-Schnittstellen mit einer wechselseitigen Bestärkung von Mensch und technischem System.

**Abbildung 1: Überblick über Cluster und Kriterien Mensch-Maschine-Interaktion im Arbeitskontext**



## 2.1 Cluster 1: Schutz des Einzelnen

Die Sicherheit Lernender Systeme und der Schutz des Einzelnen vor Risiken und negativen Folgen der Nutzung Künstlicher Intelligenz sind ein entscheidendes Element für den Einsatz und die Akzeptanz im beruflichen Umfeld, das bereits bei der Entwicklung berücksichtigt werden sollte. Neben den verschiedenen Anforderungen des Arbeits- und Gesundheitsschutzes geht es auch um die Einhaltung der Gesetze zum Schutz der Persönlichkeitsrechte der Beschäftigten und den Ausschluss von ungerechtfertigter Leistungsmessung sowie den Schutz vor Diskriminierung von Mitarbeiterinnen und Mitarbeitern. Die nachfolgenden drei Gestaltungskriterien bieten zentrale Ansatzpunkte, um den Einzelnen bei der Nutzung und Anwendung Künstlicher Intelligenz zu schützen.

## Kriterium 1: Sicherheit und Gesundheitsschutz

**Ziele:** Förderung der Gesundheit von Beschäftigten; Vermeidung von Risiken für die physische und psychische Gesundheit; Schutz vor Unfall- und Schadensereignissen (Personen- und Sachschäden) und Vermeidung von negativen physischen oder psychischen Beanspruchungsfolgen; Schutz durch Minimierung von Risiken und Fehlentscheidungen.

**Ansatzpunkte:** KI-basierte Assistenzsysteme werden in unterschiedlichen Bereichen eingesetzt – daher variieren auch die damit verbundenen prioritären Zielsetzungen im Hinblick auf Sicherheit und Gesundheitsschutz.

Bei industriellen, maschinellen Systemen steht die Vermeidung von Unfall- und Schadensereignissen (Personen- und Sachschäden) im Vordergrund. Dabei ist zu berücksichtigen, dass durch eine gute Gestaltung der Mensch-Maschine-Interaktion negative physische oder psychische Beanspruchungsfolgen (z. B. Monotonie oder psychische Sättigung) vermieden und so ein Beitrag zur Unfall- und Schadensprävention geleistet werden kann.

Bei der informations- und interaktionsbezogenen Wissens- und Dienstleistungsarbeit stehen – unter dem Aspekt der Sicherheit – vor allem der Schutz vor finanziellen und beruflichen Risiken sowie Reputationsschäden aufgrund von systemimmanenten Planungsfehlern sowie Fehlentscheidungen und Handlungen mit negativen Folgen im Mittelpunkt. In Bezug auf Gesundheitsschutz geht es hier vor allem um die Vermeidung von psychischen Fehlbeanspruchungen an der Mensch-Maschine-Schnittstelle. Bei diesen Systemen ist davon auszugehen, dass sie nicht alleine eine finale Leistung oder einen Effekt erbringen, sondern insbesondere entscheidungs- und handlungsunterstützend eingesetzt werden.

Insgesamt geht es darum, Sicherheitsaspekte mit einer menschengerechten Gestaltung von KI-Arbeitssystemen zu verknüpfen und so Risiken für die psychische Gesundheit und das Wohlbefinden von Mitarbeiterinnen und Mitarbeitern (z. B. durch überforderndes oder unerwartetes Verhalten von KI-Assistenzsystemen) zu verhindern.

## Kriterium 2: Datenschutz und verantwortungsbewusste Leistungserfassung

**Zielsetzung:** Schutz und Stärkung der Persönlichkeitsrechte der Beschäftigten; Datensparsamkeit und Zweckbindung der Datennutzung bei Künstlicher Intelligenz; rechtskonforme, verantwortungsbewusste, mitbestimmte, unternehmens- und branchenspezifische Nutzung der Möglichkeiten der Leistungserfassung; Vermeidung von Datenanalysen zur individuellen Leistungs- oder Verhaltenskontrolle; Entwicklung einer positiven, „analog“ begleiteten Kultur des Leistungsfeedbacks; Transparenz über Datenanalysen und deren Nutzung und Befähigung von Beschäftigten zum Umgang mit Datentransparenz.

**Ansatzpunkte:** Kern Lernender Systeme ist es, Daten mit ausreichender Qualität aufzunehmen und mit Hilfe verbesserter technischer Lösungen und Methoden zu analysieren und auszuwerten. Zunehmend handelt es sich dabei auch um sensible Verhaltens- und Leistungsdaten. Gleichzeitig werden die Erfassung, Speicherung und Auswertung dieser Daten immer einfacher.

Im Arbeitskontext lassen sich Datenanalysen prinzipiell als Sicherheitstechnologien, zur Verbesserung von Produktqualität und Prozessen, zur Fehlervermeidung und Assistenz sowie zu Entlastung und teils zur Prävention nutzen – aber auch zur Leistungskontrolle und Überwachung. Es macht Sinn, zwischen aggregierten und individualisierten Daten zu unterscheiden. So können beispielsweise persönliche Profile und Nutzungsdaten Arbeitsprozesse für die einzelnen Nutzerinnen und Nutzer erleichtern und auf aggregierter Stufe für das Unternehmen zur Qualitäts- und Effizienzsteigerung beitragen. Findet eine solche Trennung nicht statt, lassen sich aus analysierten persönlichen Nutzerprofilen Schlussfolgerungen ziehen, die im Extremfall zum „gläsernen“ Beschäftigten und durchgängiger individueller Leistungskontrolle führen können. Reine Verhaltens- und Leistungskontrollen der Beschäftigten sind zu vermeiden. Dies gilt insbesondere auch mit Blick auf Vorhersagen auf der Grundlage von persönlichen oder personenbezogenen Beschäftigtendaten. Zudem bergen gerade die Methoden der prädiktiven Analytik durch Künstliche Intelligenz die typischen Problematiken von Fehlschlüssen, Diskriminierung und indirekter Verhaltensbeeinflussung.

Da sowohl Datenzugang wie auch Nutzung der Analysemethoden immer einfacher werden, wird ein verantwortungsbewusster Umgang mit dem Thema Leistungserfassung zunehmend wichtig. Konkret impliziert dies den Schutz und die Stärkung der Persönlichkeitsrechte als Grundlage, speziell in Form einer Definition und differenzierten Betrachtung des Zweckes der KI-Systeme sowie der dazu nötigen Datenverarbeitung (z. B. Verbesserung der Prozesse, Weiterentwicklung der Mitarbeiterinnen und Mitarbeiter, Schutz vor Überlastung der Beschäftigten) und Vermeidung einer undifferenzierten Erfassung und Nutzung sämtlicher verfügbaren Leistungsdaten. Auch mit Blick auf die Datenqualität macht es Sinn, nur die Daten zu erheben, die für eine Anwendung tatsächlich erforderlich sind (Datensparsamkeit). Branchen- beziehungsweise unternehmensbedingte Unterschiede sind dabei zu berücksichtigen.

Zuvorderst sind bei der Entwicklung und beim Einsatz von Künstlicher Intelligenz die Gesetze zum Schutz der Persönlichkeitsrechte der Beschäftigten einzuhalten, von der Datenschutzgrundverordnung (DSGVO) über das Arbeitsrecht bis hin zu Mitbestimmungsgesetzen. So ist es etwa grundsätzlich unzulässig, biometrische Daten zur eindeutigen Identifizierung der Beschäftigten zu verarbeiten. Für die Gestaltung der Mensch-Maschine-Interaktion bedeutet das, sowohl vertrauenswürdige Verfahren und absichernde Regelungen zu implementieren als auch Datenschutzbestimmungen möglichst bereits im technischen Design der Systeme zu berücksichtigen oder auch, Datenschutz durch das Design zu bestärken. So können Kameras Gesichter bereits „on-board“ anonymisieren, noch bevor sie die Bilddaten übermitteln. Auch können Daten mit Löschfunktionen (etwa nach

Nutzung oder einer bestimmten Zeit) gekoppelt oder Zugriffsrechte und Verwendungsmöglichkeiten eingeschränkt werden.

Bereits beim Design der Systeme und dann auch beim Einsatz im Unternehmen wird also die Frage zunehmend wichtiger, wofür und wie die erfassten und analysierten Daten verwendet werden sollen. Über die technische Gestaltung der Mensch-Maschine-Interaktion zum Persönlichkeits- und Datenschutz hinaus sind flankierende Maßnahmen (z. B. Führung, Aufklärung, Aufbau von Vertrauen) wie auch ergänzende Regelungen erforderlich.

### Kriterium 3: Vielfaltssensibilität und Diskriminierungsfreiheit

**Zielsetzung:** Schutz vor Diskriminierung<sup>4</sup> von Individuen oder Gruppen und Vermeidung von möglichen Verzerrungen; bestehende Rechtsordnung als Grundlage für die Vielfaltssensibilität und Diskriminierungsfreiheit von Künstlicher Intelligenz.

**Ansatzpunkte:** Diskriminierung als ungerechtfertigte Gleich- oder Ungleichbehandlung ist von sachlich begründeten Unterscheidungen, die für erwünschte und akzeptierte Anwendungen und Funktionalitäten gerechtfertigt oder sogar erforderlich sind, zu differenzieren.

Diskriminierung kann aus unterschiedlichen Gründen bei Künstlicher Intelligenz und Lernenden Systemen entstehen beziehungsweise reproduziert und verstärkt werden: Zum einen können gesellschaftlich etablierte Vorurteile von (selbst-)lernenden Systemen – zum Beispiel durch Algorithmen oder Trainingsdaten – perpetuiert werden (prä-existierender Bias); zum anderen können technische Verfälschungen – etwa in der Sensorik – zu Diskriminierung führen (technischer Bias); darüber hinaus kann eine ungerechtfertigte Gleich- oder Ungleichbehandlung durch das Zusammenspiel zwischen Software und Anwendung erfolgen (emergenter Bias).<sup>5</sup>

Ausgangspunkt für die Einschätzung und Bewertung, ob es zu Diskriminierung in der Mensch-Maschine-Interaktion kommt, ist die bestehende Rechtsordnung. Aufbauend darauf bedarf es einer geeigneten Analyse der Umstände und Gründe. Die Vermeidung eines prä-existenten, technischen und/oder emergenten Bias erfordert Sensibilität für Vielfalt bei der Entwicklung und Anwendung der KI-Systeme ebenso wie eine sorgfältige Auswahl der Trainingsmethoden und -daten. Zudem ist ein Bewusstsein dafür nötig, dass Künstliche Intelligenz und Lernende Systeme nicht zwangsläufig neutraler oder objektiver entscheiden als Menschen (Plattform Lernende Systeme 2019).

4 Zwar ist oftmals eine „gerechtfertigte Diskriminierung“ beziehungsweise Unterscheidung (engl. discriminate; lat. discriminare) erforderlich, um korrekte Ergebnisse von KI-Systemen zu erhalten – wie etwa bei medizinischen Diagnosen durch die Unterscheidung von Frauen und Männern (vgl. Plattform Lernende Systeme 2019). Das vorliegende Papier folgt jedoch der etablierten Definition von Diskriminierung als immer sachlich unbegründete Gleich- oder Ungleichbehandlung.

5 Lernende Systeme sind umgekehrt auf einen (wohlüberlegten) Bias angewiesen, um überhaupt lernen zu können (sogenannter „learning bias“, der es ermöglicht, von gegebenen Daten auf andere zu schließen). Beim Thema Diskriminierung geht es darum, einen ethisch ungerechtfertigten beziehungsweise sozial unerwünschten Bias zu vermeiden.

## 2.2 Cluster 2: Vertrauenswürdigkeit

Damit Beschäftigte KI-Systemen vertrauen können, müssen sich Hersteller und einsetzende Unternehmen durch die konkrete menschenzentrierte Ausgestaltung der Technologie als vertrauenswürdig erweisen. Dies bedeutet, dass insbesondere bei der konkreten Erfahrung in der Mensch-Maschine-Interaktion Vertrauen schrittweise aufgebaut oder auch schlagartig reduziert werden kann. Dies betrifft sowohl das Vertrauen in das technische System als auch in die an der Herstellung und dem Einsatz beteiligten Unternehmen sowie in einzelne Personen. Die nachfolgenden drei Kriterien sind zentrale Gestaltungsfelder zur Förderung der Verlässlichkeit und in einem weiteren Schritt zur Förderung der Vertrauenswürdigkeit und Akzeptanz von Künstlicher Intelligenz in der Mensch-Maschine-Interaktion (Trustworthy AI) (EU High-Level Expert Group 2019).

### Kriterium 4: Qualität der verfügbaren Daten

**Zielsetzung:** Vermeidung von qualitativ unzureichendem Datenmaterial und entsprechenden negativen Folgen; Vorbeugung von verzerrten Datensätzen, Fehlern beziehungsweise Fehlinterpretationen und Diskriminierung; Erhöhung der Qualität statistischer Vorhersagen durch Künstliche Intelligenz; Verbesserung der Mensch-Maschine-Interaktion durch verlässliche Datengrundlagen.

**Ansatzpunkte:** Für die Sicherstellung der Datenqualität lassen sich verschiedene Anhaltspunkte benennen. Es sollten nur die Daten erhoben werden, die für eine Anwendung tatsächlich erforderlich sind. Wichtig ist dabei, bereits in der Entwicklung klare Vorstellungen über die notwendigen Daten zu entwickeln und sicherzustellen, dass ausreichend Daten für den angestrebten Zweck – auch und gerade im globalen Wettbewerb – generiert werden (Design). Erhobene Prozessdaten (nicht Personendaten oder personenbezogene Daten) in hoher Qualität können zudem auch für andere Zwecke und in anderen Kontexten genutzt werden, sodass mit bestehenden Daten gearbeitet werden kann und keine zusätzlichen Daten erhoben werden müssen (Kollateralnutzen).

Die inhaltliche Qualität der Datengrundlage ermöglicht eine gezieltere Gestaltung zuverlässiger und präziser Mensch-Technik-Schnittstellen. So kann Fehlern, Fehlschlüssen sowie Diskriminierung und entsprechenden negativen Folgen vorgebeugt und die Qualität des Interaktionsergebnisses erhöht werden. Beispielsweise wird es möglich, mittels Anwendung adaptiver Konfidenzbereiche (statistischer Erwartungsbereich) auf erfasste Datenbestände Fehlinterpretationen zu vermeiden und die Datensignifikanz und somit die Datenqualität zu erhöhen. Eine hohe Datenqualität (z. B. in Bezug auf Konsistenz, Vergleichbarkeit, Reliabilität, inhaltliche Validität) ist zudem Grundlage für sichere und adaptive Systeme, die trainiert werden, an menschliches Handeln anzuschließen, wodurch die Interaktion „humanisiert“ beziehungsweise an der menschlichen Interaktion/Kommunikation orientiert und somit menschengerechter gestaltet werden kann (Interaktionsqualität). Umgekehrt entsteht durch eine solche adaptivere Gestaltung auch eine bessere Datengrundlage in der Interaktion.

Auch die Softwareentwicklung kann zur Datenqualität beitragen, zum Beispiel indem Datenstrukturen, Datenverarbeitung und Datentransport getrennt werden. Dieses Entwicklungsprinzip („Separation of Concerns“) ist aus technischer Sicht bedeutsam, hat aber gleichzeitig auch Auswirkung auf andere Bereiche – wie beispielsweise Sicherheit oder Transparenz.

### Kriterium 5: Transparenz, Erklärbarkeit und Widerspruchsfreiheit

**Zielsetzung:** Umsetzung von Ansätzen einer erklärbaren Künstlichen Intelligenz (Explainable AI); Entwicklung von Möglichkeiten zur Nachvollziehbarkeit von Lernenden Systemen und Herstellung (abgestufter) Transparenz über deren Entscheidungsprozesse und Entscheidungen; Verhinderung von Demotivation und Überbeanspruchung von Beschäftigten durch eine widerspruchsfreie Gestaltung der Mensch-Maschine-Interaktion.

**Ansatzpunkte:** In der Mensch-Maschine-Interaktion sind die Menschen vor eine für sie oft nicht nachvollziehbare Komplexität des ihnen gegenüberstehenden Systems gestellt. Ein Mangel an Nachvollziehbarkeit kann zu Demotivation und zu einer Ablehnung solcher Systeme führen. Um dem entgegenzuwirken, müssen Lernende Systeme so gestaltet sein, dass sie den Beschäftigten Basisinformationen über ihre prinzipielle Funktionsweise, ihre eingeschriebenen Zwecke und Zielsetzungen, ihren Datenfokus und Datengrundlage und die davon abhängigen neu gebildeten Kategorien, „Hypothesen“, Ergebnisse und vor allem Folgerungen und Entscheidungen beziehungsweise Empfehlungen (Output) liefern. Hierfür sind niedrigschwellige, zielgruppenorientierte, lern- und erfahrungsförderliche Lösungen gefordert. Dabei kann auch je nach Aufgabe und Rolle der Nutzerin und des Nutzers sowie je nach Anwendungsgebiet eine Abstufung der Transparenz erfolgen.

Alle für die Interaktion relevanten Informationen müssen so aufbereitet sein, dass sie für die betroffenen Menschen verständlich sind. Entsprechende Anhaltspunkte zur Lösung dieses Problems bieten die Ansätze aus dem Bereich der erklärbaren Künstlichen Intelligenz (Explainable AI), die Methoden zu entwickeln versucht, die für die Nachvollziehbarkeit Lernender Systeme dienlich sind. Dabei stehen insbesondere die Selbstbeschreibungsfähigkeit und die Erwartungskonformität im Mittelpunkt. Dadurch leistet die erklärbare Künstliche Intelligenz einen wichtigen Beitrag zur Entwicklung einer verantwortungsbewussten Künstlichen Intelligenz (Responsible AI), die sich durch Transparenz, Fairness, Verlässlichkeit und die Orientierung an ethischen Vorstellungen auszeichnet.

Ein wesentlicher Aspekt in Bezug auf die kognitive und soziale Belastung von Beschäftigten ist zudem die konsistente und widerspruchsfreie Gestaltung der Interaktion mit dem Lernenden System. Dabei sind zum einen widersprüchliche Informationen und Prozesse bei der direkten Interaktion an der Schnittstelle zu vermeiden, die Frustration erzeugen. Zum anderen muss aber auch besonders darauf geachtet werden, keine betrieblichen Zielkonflikte in die KI-Systeme zu verlagern – wie etwa häufig anzutreffende betriebliche

Widersprüche und Anforderungen im Dreieck zwischen Zeitaufwand, Kosteneffizienz und Qualität oder auch typische Spannungen zwischen abstrakter Planung (Soll-Situation) und konkreter Umsetzung (Ist-Situation).

Werden diese Widersprüche und Spannungen in die technischen Arbeitssysteme eingeschrieben und damit „objektiviert“ und nicht sozial vermittelt oder ausgehandelt, besteht die Gefahr, dass sie die Handlungsräume und Selbstwirksamkeit der Beschäftigten verringern und so zu negativen Beanspruchungsfolgen führen können. Insgesamt sollten Nutzerinnen und Nutzer auch in der Interaktion mit Künstlicher Intelligenz nur vor bewältigbare Situationen gestellt werden. Bereits beim Design ist deshalb darauf zu achten, dass die Interaktion nicht zu negativen psychischen Beanspruchungsfolgen führt (was im Arbeitsprozess etwa mittels gesetzlich vorgeschriebener Gefährdungsbeurteilungen zu überprüfen ist).

Werden Lernende Systeme in der Mensch-Maschine-Interaktion als nachvollziehbar, widerspruchsfrei und angemessen sowie als Teil eines verantwortungsbewussten Systems gestaltet, bildet dies eine stabile Grundlage für Vertrauen und letztlich auch Akzeptanz Lernender Systeme.

### Kriterium 6: Verantwortung, Haftung und Systemvertrauen

**Zielsetzung:** Zuschreibbarkeit von Verantwortung; Kompetenz und Ressourcen sowie Kontrolle über das System als Voraussetzung für die Übernahme von Verantwortung; Ausmaß der Systemkontrolle als Gradmesser für Art und Umfang der Verantwortung von Nutzerinnen und Nutzern; Ansatzpunkte für die Gestaltung durch Konzepte der vertrauenswürdigen Künstlichen Intelligenz (Trustworthy AI).

**Ansatzpunkte:** Notwendige Voraussetzungen für die Übernahme von Verantwortung für Unternehmen wie auch Beschäftigte sind Transparenz (einschließlich der Zielsetzungen) und Beherrschbarkeit Lernender Systeme. Dazu gehört auch, dass explizit definiert wird, welche Informationen, Ressourcen und Kompetenzen gegeben sein müssen, damit Menschen mit ihnen interaktiv handlungsfähig sind. Wichtig ist zudem, dass die Mensch-Maschine-Interaktion so ausgelegt ist, dass sie es den Beschäftigten erlaubt, nach ihrer Arbeitsorientierung und den geltenden Regeln zu handeln und gegebenenfalls eine Interaktion auch abzubrechen. Nur so sind Nutzerinnen und Nutzer bereit und in der Lage, Verantwortung für ihren Part in der Interaktion im soziotechnischen System wahrzunehmen. Ebenso sind Transparenz und Beherrschbarkeit erforderlich, damit Lernende Systeme als vertrauenswürdig wahrgenommen werden.

Das Abschätzen eines zukünftigen Systemverhaltens ist die Basis für die menschliche Kontrolle des technischen Systems. Die menschliche Interaktion mit dem System sollte daher auf einem adäquaten Abstraktionsniveau stattfinden und im Bereich des Erwartbaren liegen. Dieser Wissens- und Handlungsraum muss auf die Interaktion zugeschnitten im Vorfeld (möglichst partizipativ) bestimmt- und abgestimmt werden. Gerade in Bezug auf

die Verankerung von Vertrauen, Verantwortung und Haftung sind dabei bereits im Systemdesign ausreichend Tests und eine Technikfolgenabschätzung zu absolvieren, um unerwünschtes Systemverhalten auszuschließen. Die Möglichkeit der Systemkontrolle durch Unternehmen und Beschäftigte ist sicherzustellen, dabei werden Verantwortlichkeiten festgelegt. Der Grad bestimmt zudem wesentlich die Art und das Maß von Verantwortung. In der Praxis, wie auch im grundsätzlichen juristischen Sinne, ist die Möglichkeit der Kontrolle mit der Verortung von Verantwortung bei der jeweiligen juristischen Person verknüpft. Gleichzeitig wird dadurch auch das menschliche Erleben von Selbstwirksamkeit nachhaltig unterstrichen.

Zudem muss das Systemdesign so angelegt sein, dass Prozesse (Input, Verarbeitung und Output) auch ex-post nachvollzogen und Ursachen für eventuelle Problematiken identifiziert werden können. Hier müssen entsprechende Methoden bereits im Design vorgesehen sein, die dies trotz hoher Komplexität leisten können.

Das Vertrauen in eine Technologie ist ein individueller und sozial konstruktiver Prozess, bei dem sich aufgrund von Erfahrungen und Wissen die Qualität des Vertrauens und auch die Kriterien der Bewertung wandeln können. Diese Anforderung lässt sich durch die Aspekte Verstehbarkeit, Bewältigbarkeit und Sinnhaftigkeit des Erlebens einer Situation und Interaktion mit einem (selbst-)lernenden System bereits beim Design konkretisieren. Wichtige Ansatzpunkte bietet das Konzept der vertrauenswürdigen Künstlichen Intelligenz (Trustworthy AI), das von einer durch die Kommission der Europäischen Union eingesetzten Hochrangigen Expertengruppe für Künstliche Intelligenz erarbeitet wurde und durch Ethik-Leitlinien einen Rahmen für den Einsatz Künstlicher Intelligenz formuliert (EU High-Level Expert Group 2019).

### **2.3 Cluster 3: Sinnvolle Arbeitsteilung**

Der Einsatz von KI-Technologien und Lernenden Systemen wirkt sich nicht nur auf Tätigkeitsprofile, Kompetenzanforderungen und Arbeitsprozesse aus, sondern betrifft die „Arbeitsteilung“ von Mensch und Maschine insgesamt. Dabei gilt es, mit einer menschenzentrierten Gestaltung der Mensch-Maschine-Interaktion eine sinnvolle „Arbeitsteilung“ zu finden sowie Beschäftigte in ihrer Arbeit nachhaltig zu entlasten und zu unterstützen. Gleichzeitig stehen dabei Fragen der Handlungsträgerschaft und der Situationskontrolle sowie Themen der flexiblen und situationsspezifischen Anpassung der Systeme im Vordergrund. Die nachfolgenden drei Kriterien adressieren die verschiedenen Anforderungen an eine sinnvolle „Arbeitsteilung“ zwischen Mensch und Maschine. Dabei ist zu beachten, dass die jeweilige Ausgestaltung der Kriterien auf die Aufgaben sowie auf das Qualifikationsprofil und die Kompetenzen der Mitarbeiterinnen und Mitarbeiter zugeschnitten sein sollte.

## Kriterium 7: Angemessenheit, Entlastung und Unterstützung

**Zielsetzung:** Komplementäre Ergänzung der Fähigkeiten von Menschen und Maschinen für eine sinnvolle „Arbeitsteilung“; angemessene Arbeitsinhalte und Anforderungen; nachhaltige Entlastung und Unterstützung von Beschäftigten durch KI-basierte Assistenzsysteme.

**Ansatzpunkte:** Die Nutzung von Lernenden Systemen in unterschiedlichen Branchen und Bereichen führt zu einer neuen Aufgabenverteilung zwischen Mensch und Technik. Entscheidend ist dabei die Berücksichtigung der unterschiedlichen Fähigkeiten und Eigenschaften von Mensch und Technik bereits beim Design der interaktiven Systeme. Durch eine komplementäre Ergänzung der jeweiligen spezifischen Stärken können ein wechselseitiges Bestärkungsverhältnis geschaffen, negative Beanspruchungsfolgen vermieden sowie eine Entlastung und Unterstützung von Beschäftigten erreicht werden. Auf diese Weise bietet Künstliche Intelligenz auch die Chance auf eine Förderung einer positiven individuellen Auslastung bei gleichzeitiger Vermeidung von Überlastung oder Überforderung.

Wichtig ist es, bei der Entwicklung und bei der Implementierung von interaktiven KI-Systemen darauf zu achten, dass die Interaktion auf die Qualifikationen und Kompetenzen der Nutzerinnen und Nutzer zugeschnitten gestaltet wird – sowohl in Bezug auf die Inhalte als auch im Hinblick auf die Form der Interaktion (z. B. Reaktionszeiten oder Arbeitsdichte). Notwendig ist dabei eine situations- wie nutzerspezifische Gestaltung bereits beim Design: Die Technik soll dem Menschen angepasst werden. Gleichzeitig müssen Beschäftigte befähigt werden, mit KI-Systemen „zusammenarbeiten“ zu können. Dies sollte bereits durch die technische Gestaltung gefördert werden (siehe Kriterium 1.1). Nicht nur für die Beschäftigten, sondern auch für Führung und Management ist es zentral, die Potentiale, aber auch die Grenzen Künstlicher Intelligenz und Lernender Systeme realistisch einschätzen zu können, um diese sinnvoll zu nutzen und Fehleinschätzungen zu reduzieren.

Zudem soll der Mensch in seiner zentralen Rolle im Arbeitsprozess nicht geschwächt, sondern durch die Interaktion mit Künstlicher Intelligenz gestärkt werden. KI-basierte Assistenzsysteme sollten daher zum einen auf eine Entlastung der Beschäftigten von (physisch oder psychisch) anstrengenden oder gefährlichen Tätigkeiten ausgerichtet sein. Zum anderen können Lernende Systeme den Menschen bei komplexen Sachverhalten und Aufgaben unterstützen – etwa in schwierigen Entscheidungssituationen – und so ermächtigen und befähigen (Empowerment).

## Kriterium 8: Handlungsträgerschaft und Situationskontrolle

**Zielsetzung:** Gezielte und transparente Gestaltung der Handlungsträgerschaft und Situationskontrolle in der Mensch-Maschine-Interaktion; Minimierung und Vermeidung von Risiken und negativen Beanspruchungsfolgen.

**Ansatzpunkte:** Eine zentrale Forderung der humanen Arbeitsgestaltung ist, dass Menschen nicht gezwungen werden sollen, ihr Handeln an technischen Systemen auszurichten. Bei der Interaktion mit Lernenden Systemen ist jedoch – als typisches Kennzeichen komplexer sozio-technischer Arrangements – ein Teil der Handlungsträgerschaft beziehungsweise Situationskontrolle (Wer stößt Handlungen an? Wer koordiniert eine Situation beziehungsweise Interaktionssequenz?) auf Seite des technischen Systems eingeschrieben. Interaktive KI-Systeme können so entwickelt sein, dass sie ihre Nutzerinnen und Nutzer zu einem vorherbestimmten und teilweise auch zu einem neu hervorgebrachten Handeln „auffordern“ beziehungsweise „nötigen“, sodass das Lernende System dem Menschen in bestimmten Situationen als eine Art „Akteur“ gegenübertritt. Aus der Praxis im Rahmen der Mensch-Roboter-Kollaboration zeigt sich, dass den Beschäftigten vor Ort Selbstwirksamkeit wichtig ist. Praktisch bedeutet dies, dass ein hohes Maß an Handlungsträgerschaft und Situationskontrolle Unzufriedenheit bei den Mitarbeiterinnen und Mitarbeitern vorbeugen kann.

Wenn Mensch und Technik hochgradig interaktiv und komplementär „arbeitsteilig“ kollaborieren und voneinander lernen sollen, muss die Handlungsträgerschaft eindeutig geklärt werden. Dabei ist es wichtig, dass klar ist, wo die Handlungsträgerschaft zwischen KI-Anwendung und Beschäftigten zu einem bestimmten Zeitpunkt liegt und wann konkrete Übergaben in der Interaktion erfolgen. Es muss nachvollziehbar sein, wer gerade was zu einem gemeinsamen Prozess beiträgt, bei wem die Situationskontrolle für den jeweiligen Teilprozess liegt und wo oder wie die Teilprozesse aneinander anschließen. Dabei ist es erforderlich, Regeln zu definieren und Möglichkeiten zu schaffen, wie Übergaben zwischen Mensch und Maschine (KI-Anwendung) wechselseitig gezielt und transparent angestoßen werden können – zum Beispiel, wenn das System überfordert ist, der Mensch intervenieren will oder Unterstützung anfordert.

Dies ist die Ausgangsbasis für eine potentielle Verantwortungszuschreibung und eine entsprechende Entlastung der Situation von unklaren Risiken. Zudem werden so Grundlagen für eine transparente und nachvollziehbare Gestaltung Lernender Systeme formuliert, Möglichkeiten für menschliche Interventionen und betriebliche Regelungen eröffnet sowie Optionen für eine lernförderliche Gestaltung der Mensch-Maschine-Interaktion ermöglicht. Nicht zuletzt ist eine transparente und interaktiv abstimmbare Klärung der Handlungsträgerschaft Dreh- und Angelpunkt für eine komplementäre „Arbeitsteilung“ zwischen Mensch und KI-System.

## Kriterium 9: Adaptivität, Fehlertoleranz und Individualisierbarkeit

**Zielsetzung:** Befähigung Lernender Systeme, sich flexibel und situationspezifisch an den Bedarfen und Bedürfnissen sowie an der Arbeitspraxis der Nutzerinnen und Nutzer ausrichten zu können.

**Ansatzpunkte:** Mit Lernenden Systemen ist die Hoffnung verbunden, dass sie sich flexibler an Veränderungen ihrer Umwelt anpassen und Komplexität bearbeiten können beziehungsweise eine hohe Adaptivität aufweisen. Besonders eine im Sinne der Nutzerinnen und Nutzer nachhaltige Gestaltung der Mensch-Maschine-Interaktion erfordert eine hohe soziale Adaptivität des KI-Systems.

Dies bedeutet, dass Lernende Systeme nicht nur fähig sein müssen, Anforderungen aus der Umwelt in die eigene Systemlogik zu übersetzen („assimilierende Adaptivität“), sondern die eigene Bearbeitungslogik auch an die Bedarfe der Umwelt und vor allem der Beschäftigten anzupassen („komplementäre Adaptivität“).

Für stark interaktive KI-Systeme im Arbeitszusammenhang ist dies höchst voraussetzungsreich. Denn dies bedeutet zum Beispiel, dass bei ihrem Design „eingeplant“ werden muss, dass KI-Systeme unvorhergesehene Aneignungsprozesse in der Nutzung zulassen und sogar unterstützen. Dies geht weit über übliche Anforderungen an Technikentwicklung hinaus – wie etwa Robustheit, Fehlertoleranz und Individualisierbarkeit.

Letzten Endes verweist dieses Kriterium auf die neue Möglichkeit Lernender Systeme, eine implizite Programmierbarkeit im Nutzungsprozess zu ermöglichen. Damit stünde nicht mehr das Nutzererlebnis, sondern die Nutzerermächtigung im Mittelpunkt des Designs.

## 2.4 Cluster 4: Förderliche Arbeitsbedingungen

Lernende Systeme können menschliche Arbeit übernehmen und damit die Gelegenheiten sinnlicher Wahrnehmung, des Lernens und Einübens sowie Kompetenzerlebnisse potentiell verringern. Bei der Gestaltung Künstlicher Intelligenz sollte daher darauf geachtet werden, menschliche Grundbedürfnisse gezielt zu adressieren. Das gilt sowohl mit Blick auf Aufgabenprofile und die „Arbeitsteilung“ von Mensch und Maschine als auch für das Design der Benutzungsschnittstelle. Dies bedeutet konkret, dass Handlungsräume und reichhaltige Arbeit gesichert, wechselseitige Lern- und Erfahrungsförderlichkeit für Mensch und KI-System ermöglicht sowie die Mensch-Maschine-Interaktion sensibel für die Anforderungen und Funktionen von Kommunikation, Kooperation und Einbindung zu gestalten sind. Im Folgenden werden drei weitere Gestaltungskriterien vorgestellt.

## Kriterium 10: Handlungsräume und reichhaltige Arbeit

**Zielsetzung:** Sicherung und (wo sinnvoll auch) Ausweitung von Handlungsspielräumen von Beschäftigten (insbesondere Autonomie und Entscheidungsfreiheit sowie Vielfalt an Handlungsmöglichkeiten); Einbeziehung von menschlichen Grundbedürfnissen für eine sinnvolle, motivierende und gesundheits- sowie persönlichkeitsförderliche Arbeit.

**Ansatzpunkte:** Als Handlungsspielraum wird der Grad an Autonomie und Entscheidungsfreiheit sowie die Vielfalt an Möglichkeiten im Handeln und Gestalten einer Person in der Arbeit verstanden. Dies betrifft sowohl die Ziele, die Arbeitsinhalte und die konkrete Ausführung als auch die strukturierenden (organisationalen und technischen) Rahmenbedingungen. Ergänzend kommt der Aspekt der reichhaltigen Arbeit, also die Anreicherung von Tätigkeiten mit abwechslungsreichen, herausfordernden und fördernden Inhalten (Job-Enrichment), hinzu. Die Erweiterung von Handlungsspielräumen und reichhaltige Arbeit sind Kernpunkte der Humanisierung der Arbeitswelt (HdA) und bieten wichtige Anhaltspunkte für die Gestaltung der Mensch-Maschine-Interaktion.

Für die Gestaltung und den Einsatz von Künstlicher Intelligenz bedeutet das, dass darauf geachtet werden muss, dass diese technischen Systeme die Handlungsräume der Nutzerinnen und Nutzer nicht einengen und nicht eben jene Arbeitsinhalte übernehmen, die in der Arbeit motivierend, qualifizierend und gesundheitsförderlich wirken. Künstliche Intelligenz sollte im Gegenteil befähigen und Handlungsräume erweitern, indem vorher unmögliche Handlungen erreichbar werden. Auch die Mensch-Maschine-Interaktion bei Künstlicher Intelligenz kann selbst neue herausfordernde, motivierende und entwicklungsförderliche Arbeitsinhalte bereitstellen.

## Kriterium 11: Lern- und Erfahrungsförderlichkeit

**Zielsetzung:** Mensch-Maschine-Interaktion lern- und erfahrungsförderlich gestalten; Ermöglichung des wechselseitigen Lernens des Menschen von der Maschine und umgekehrt; nachvollziehbare und adaptive Gestaltung der Systeme zur Aneignung und Integration von Wissen und Erfahrung; Gewährleistung der Überführung von Daten in Information beziehungsweise von Information in Wissen.

**Ansatzpunkte:** Voraussetzung für die Umsetzung des Kriteriums ist die Berücksichtigung der essenziellen Unterschiede zwischen Mensch und Maschine bei Erwerb, Verarbeitung und Speicherung, Reproduktion und Abrufen sowie Anwenden von Wissen, um den heterogenen Anforderungen beider Seiten an das Lernen gerecht zu werden und wechselseitiges Lernen zu ermöglichen.

Daher muss erstens die Interaktion mit KI-Systemen für die Nutzerinnen und Nutzer lern- und erfahrungsförderlich gestaltet werden. Dies betrifft die nachvollziehbare (Explainable AI) und wechselseitig adaptive Gestaltung, um die Aneignung von Wissen und Erfahrung im Nutzungsprozess zu ermöglichen. An dieser Stelle ergeben sich mögliche Zielkonflikte

etwa in Bezug auf Beschäftigendaten, um zum Beispiel individualisiertes Lernen zu ermöglichen, welche sich jedoch durch geeignete Regelungen und ein entsprechendes Design lösen lassen. Zudem geht es auch um die gezielte Integration von expliziten Lerninhalten (Qualifikationen) in die Mensch-Maschine-Interaktion und eine entsprechend didaktische Gestaltung zum Beispiel bei Assistenzsystemen.

Zweitens kann durch eine wechselseitig lernförderliche Gestaltung der Mensch-Maschine-Interaktion die Passgenauigkeit und Leistungsfähigkeit des KI-Systems verbessert werden, indem die Nutzerinnen und Nutzer befähigt werden, die Lerninhalte (Datenqualität) und das Lernverhalten (Verknüpfungen) des intelligenten Systems interaktiv zu validieren und gegebenenfalls zu korrigieren. Eine wechselseitig lernförderliche Gestaltung erhöht zudem die Wahrscheinlichkeit, dass Menschen bereit sind, ihr Wissen und ihre Erfahrungen in KI-Systeme einzubringen.

Ein solcher komplementärer Ansatz bietet insbesondere für die Bearbeitung komplexer Situationen große Chancen. Denn erst im sozio-technischen System werden Daten zu Informationen kontextualisiert und in auf konkrete Situationen anwendbares und übertragbares Wissen überführbar. Damit erst werden maschinell gelernte Inhalte und menschliches Erfahrungswissen sinnvoll integrierbar.

Gerade wenn KI-Systeme weitreichende und relevante Tätigkeiten übernehmen und Gelegenheiten für den Wissens- und Erfahrungserwerb möglicherweise knapp werden, ist eine lern- und erfahrungsförderliche Gestaltung der Mensch-Maschine-Interaktion von großer Bedeutung – vor allem um Wissen, Erfahrungen und Kompetenzen zu erhalten und zu erweitern sowie Nutzerinnen und Nutzer in ihrer Selbstwirksamkeit zu stärken; aber natürlich auch, um Innovation aus den Prozessen heraus zu fördern und die Nutzerinnen und Nutzer zu befähigen, die Performanz des KI-Systems einschätzen und gegebenenfalls berichtigen oder verbessern zu können.

### **Kriterium 12: Kommunikation, Kollaboration und soziale Einbindung**

**Zielsetzung:** Doppelte Sensibilisierung von Künstlicher Intelligenz für soziale Kontexte und Strukturen; Stärkung und Unterstützung zwischenmenschlicher Kommunikation, Kollaboration und Verbundenheit durch KI-Systeme.

**Ansatzpunkte:** Kommunikation, Kooperation und Einbindung stellen wesentliche Grundbedingungen für eine qualitativ hochwertige und effiziente Arbeit, Sinnstiftung und Integration dar. Dies betrifft einerseits die formalen Anteile von Arbeit, die im Voraus geplant und gestaltet werden und sich in formellen Arbeitsanforderungen, der Arbeitsorganisation und der Technikgestaltung widerspiegeln. Zum anderen geht es um die informelle Arbeitspraxis, situative Abstimmungen, Innovationen zur flexiblen Bewältigung der nicht planbaren Anteile von Arbeit.

Eine wesentliche Herausforderung für die Gestaltung der Mensch-Maschine-Interaktion liegt darin, Künstliche Intelligenz im doppelten Sinne „sensibel“ für soziale Kontexte und Strukturen zu machen: Zum einen kann – in sehr beschränktem Maße – Künstliche Intelligenz als „Kooperationspartner“ agieren; zum anderen muss die Mensch-Maschine-Interaktion so gestaltet sein, dass sie notwendige und gewinnbringende zwischenmenschliche Kommunikation, Kooperation und Verbundenheit nicht verhindert oder ersetzt, sondern im besten Falle sogar noch zielführend unterstützt.

KI-basierte Technologien können dabei vielfältig unterstützen – zum Beispiel, indem sie die Beschäftigten von Standardkommunikation entlasten, Daten selbstständig aus Datensilos sammeln und für Entscheidungen zur Verfügung stellen, situationsbezogenen Wissensträger „matchen“ beziehungsweise identifizieren und in Verbindung bringen oder selbst zum „Kooperationspartner“ werden. KI-basierte Technologien in interaktiven Systemen sollten derart ausgelegt werden, dass in der Kollaboration die Fähigkeiten und Fertigkeiten beziehungsweise die Kompetenzen des Menschen vom technischen System erkannt beziehungsweise aufgenommen werden und die Funktionalitäten in der Interaktion darauf adaptiert werden. So sollten zum Beispiel manuell kollaborativ agierende Roboter (Cobots) auf die physischen Fähigkeiten und Fertigkeiten sowie die Erfahrungswerte (z. B. in Bezug auf die Prozessgestaltung) abgestimmt werden.

### 3. Umsetzung der Kriterien und Ausblick

---

Die vorgestellten Kriterien sollen einen wichtigen Impuls für die nachhaltige, menschenzentrierte und zukunftsorientierte Gestaltung der Mensch-Maschine-Interaktion bei Künstlicher Intelligenz geben. Zudem sollen sie einen Beitrag zur Entwicklung einer neuen Aufgabenverteilung zwischen Beschäftigten und technischen Systemen – in diesem Falle insbesondere Lernenden Systemen – leisten.

Die Kriterien sind als eine Handreichung mit konkreten Ansatzpunkten für die Entwicklung und den Einsatz Künstlicher Intelligenz zu verstehen (zusammenfassender Überblick in Tabelle 1). Zu beachten ist dabei, dass die Gestaltungskriterien einer weiteren Konkretisierung auf den jeweiligen Anwendungsfall bedürfen. Dabei ergeben sich Ziel- und Interessenskonflikte, die im Rahmen bestehenden Rechts ausgehandelt und bearbeitet beziehungsweise geregelt werden müssen. Darüber hinaus lassen sich die Kriterien mit bestehenden ähnlichen Konzepten (wie etwa den Ansatzpunkten des Human Centered Designs oder der erfahrungsförderlichen Technikgestaltung) wechselseitig verbinden und sollen in laufende Aktivitäten – wie etwa in der Normung und Standardisierung – eingebracht werden.

Der Kriterienkatalog ist einer von drei Bausteinen eines (in Entstehung befindlichen) Gestaltungsinstruments für die Mensch-Maschine-Interaktion bei Künstlicher Intelligenz (Abbildung 2). Die Gestaltungskriterien (1) ermöglichen einen umfassenden Blick auf die Gestaltungsanforderungen. Um jedoch eine Gesamtbewertung zu ermöglichen, wird querliegend eine Bewertungsfolie (2) des gesamten sozio-technischen Arrangements eingezogen. Dabei geht es um eine Bewertung der Qualität und Intensität der Interaktion zwischen Mensch und Technik – angelehnt an unterschiedliche Autonomiestufen (acatech/ Fachforum autonome Systeme 2017, Plattform Industrie 4.0 2019a) oder verschiedene Kritikalitätsstufen.

**Abbildung 2: Gestaltungskonzept für die Mensch-Maschine-Interaktion**



Quelle: Eigene Darstellung Huchler.

Von der Bewertung der Mensch-Maschine-Interaktion aus kann der Betrachtungshorizont dann mittels einer Technikfolgenabschätzung (3) schrittweise auf die unterschiedlich weiten Folgewirkungen und Anforderungen erweitert werden: von den potentiellen Auswirkungen auf die konkrete Interaktionssituation an der Mensch-Maschine-Interaktion vor Ort, über die Folgen für das Arbeitsumfeld (wie etwa Zusammenarbeit, Führung oder Arbeitsorganisation) oder den Betrieb (wie etwa Qualifizierung, Technikentwicklung und Technikeinsatz, Produktqualität oder Wertschöpfungskonzepte) bis hin zur Gesellschaft (wie etwa Arbeitsplätze, Einkommen, Bildung, Zusammenhalt oder Verbraucherschutz).

Nicht zuletzt ist die Gestaltung der Mensch-Maschine-Interaktion bei stark interaktiven KI-Systemen nur ein Ausschnitt der Anforderungen, die Künstliche Intelligenz an den Wandel von Arbeit stellt. Die Gestaltungskriterien sollen daher durch ein weiteres Whitepaper zu Fragen der Umsetzung beziehungsweise des Change-Managements in den Unternehmen angesichts des zunehmenden Einsatzes Lernender Systeme im Arbeitsumfeld ergänzt und fortgeschrieben werden.

**Tabelle 1: Kriterien für die Gestaltung der Mensch-Maschine-Interaktion**

Schutz des Einzelnen	
Sicherheit und Gesundheitsschutz	<ul style="list-style-type: none"> <li>• Vermeidung von Risiken für die physische und psychische Gesundheit von Beschäftigten</li> <li>• Schutz vor Unfall- und Schadensereignissen (Personen- und Sachschäden)</li> <li>• Vorbeugung gegen negative physische oder psychische Beanspruchungsfolgen</li> </ul>
Datenschutz und verantwortungsbewusste Leistungserfassung	<ul style="list-style-type: none"> <li>• Schutz der Persönlichkeitsrechte, Datensparsamkeit und Zweckbindung der Datennutzung</li> <li>• Vermeidung von Datenanalysen zur ungerechtfertigten Leistungsmessung</li> <li>• Entwicklung einer positiven Kultur des Leistungsfeedbacks</li> <li>• Transparenz über und Befähigung zu Datenanalysen und deren Nutzung</li> </ul>
Vielfaltssensibilität und Diskriminierungsfreiheit	<ul style="list-style-type: none"> <li>• Schutz vor Diskriminierung von Individuen oder Gruppen</li> <li>• Bestehende Rechtsordnung als Grundlage für die Vielfaltssensibilität und Diskriminierungsfreiheit</li> </ul>

<b>Vertrauenswürdigkeit</b>	
Qualität der verfügbaren Daten	<ul style="list-style-type: none"> <li>• Vermeidung von qualitativ unzureichendem Datenmaterial</li> <li>• Vorbeugung von verzerrten Datensätzen, Fehlern/Fehlinterpretationen und Diskriminierung</li> <li>• Verbesserung der Mensch-Maschine-Interaktion durch verlässliche Datengrundlagen</li> </ul>
Transparenz, Erklärbarkeit und Widerspruchsfreiheit	<ul style="list-style-type: none"> <li>• Umsetzung von Ansätzen einer erklärbaren Künstlichen Intelligenz (Explainable AI)</li> <li>• Entwicklung von Methoden zur Nachvollziehbarkeit Lernender Systeme</li> <li>• Herstellung (abgestufter) Transparenz über Entscheidungsprozesse Lernender Systeme</li> <li>• Verhinderung von Demotivation durch widerspruchsfreie Gestaltung der Mensch-Maschine-Interaktion</li> </ul>
Verantwortung, Haftung und Systemvertrauen	<ul style="list-style-type: none"> <li>• Transparenz über und Zuschreibbarkeit von Verantwortung</li> <li>• Kompetenz und Kontrolle über das System als Voraussetzung für die Übernahme von Verantwortung</li> <li>• Ausmaß der Systemkontrolle als Gradmesser für Art und Umfang der Verantwortung</li> <li>• Ansatzpunkte entlang der Konzepte einer vertrauenswürdigen Künstlichen Intelligenz (Trustworthy AI)</li> </ul>
<b>Sinnvolle Arbeitsteilung</b>	
Angemessenheit, Entlastung und Unterstützung	<ul style="list-style-type: none"> <li>• Angemessene Arbeitsinhalte und Anforderungen</li> <li>• Komplementäre Ergänzung der Fähigkeiten von Menschen und Maschinen für eine sinnvolle Arbeitsteilung</li> <li>• Nachhaltige Entlastung und Unterstützung von Beschäftigten durch KI-basierte Assistenzsysteme</li> <li>• Befähigung zur Arbeit mit KI-Systemen</li> </ul>
Handlungsträgerschaft und Situationskontrolle	<ul style="list-style-type: none"> <li>• Gezielte und transparente Gestaltung der Handlungsträgerschaft und Situationskontrolle</li> <li>• Minimierung und Vermeidung von Risiken und negativen Beanspruchungsfolgen</li> </ul>
Adaptivität, Fehlertoleranz und Individualisierbarkeit	<ul style="list-style-type: none"> <li>• Befähigung Lernender Systeme, sich flexibel und situationsspezifisch an den Bedarfen und Bedürfnissen sowie an der Arbeitspraxis der Nutzerinnen und Nutzer ausrichten zu können</li> </ul>
<b>Förderliche Arbeitsbedingungen</b>	
Handlungsräume und reichhaltige Arbeit	<ul style="list-style-type: none"> <li>• Sicherung und gegebenenfalls Ausweitung von Handlungsspielräumen von Beschäftigten (insbesondere Autonomie und Entscheidungsfreiheit sowie Vielfalt an Handlungsmöglichkeiten)</li> <li>• Einbeziehung von menschlichen Grundbedürfnissen für eine sinnvolle, motivierende und gesundheits- sowie persönlichkeitsförderliche Arbeit</li> </ul>
Lern- und Erfahrungsförderlichkeit	<ul style="list-style-type: none"> <li>• Ermöglichung des wechselseitigen Lernens des Menschen von der Maschine und umgekehrt</li> <li>• Nachvollziehbare, adaptive Gestaltung der Systeme zur Integration von Wissen und Erfahrung</li> <li>• Gewährleistung der Überführung von Daten in Information beziehungsweise von Information in Wissen</li> </ul>
Kommunikation, Kooperation und soziale Einbindung	<ul style="list-style-type: none"> <li>• Doppelte Sensibilisierung von Künstlicher Intelligenz für die sozialen Kontexte und Strukturen</li> <li>• Unterstützung zwischenmenschlicher Kommunikation, Kooperation und Verbundenheit</li> <li>• Künstliche Intelligenz als Kooperationspartner</li> </ul>

## Über dieses Whitepaper

---

Dieses Papier wurde erstellt von der Arbeitsgruppe Arbeit/Qualifikation, Mensch-Maschine-Interaktion der Plattform Lernende Systeme. Als eine von insgesamt sieben Arbeitsgruppen untersucht sie die Potentiale und Herausforderungen, die sich aus dem Einsatz Künstlicher Intelligenz in der Arbeits- und Lebenswelt ergeben. Dabei stehen die Fragen der Transformation und der Entwicklung menschengerechter Arbeitsbedingungen im Fokus. Zudem nimmt sie die Anforderungen und Optionen für die Qualifizierung und das lebensbegleitende Lernen sowie Ansatzpunkte für die Gestaltung der Mensch-Maschine-Interaktion und die Arbeitsteilung von Mensch und Technik in den Blick.

### **Autoren:**

**Dr. Norbert Huchler** (Federführung), Institut für Sozialwissenschaftliche Forschung e. V. (ISF-München)

**Prof. Dr. Lars Adolph**, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)

**Prof. Dr. Elisabeth André**, Universität Augsburg

**Prof. Dr.-Ing. Prof. e. h. Wilhelm Bauer**, Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO und Universität Stuttgart

**Nadine Bender**, KUKA Deutschland GmbH

**Dr. Nadine Müller**, Vereinte Dienstleistungsgewerkschaft (ver.di)

**Dr. Rahild Neuburger**, Ludwig-Maximilians-Universität München

**Dr.-Ing. Matthias Peissner**, Fraunhofer-Institut für Arbeitswirtschaft und Organisation (IAO)

**Prof. Dr. Jochen Steil**, Technische Universität Braunschweig

**Prof. Dr. Ing. Sascha Stowasser**, Institut für angewandte Arbeitswissenschaft (ifaa)

**Oliver Suchy**, Deutscher Gewerkschaftsbund (DGB)

### **Die Arbeitsgruppe wird geleitet von:**

**Prof. Dr. Elisabeth André**, Universität Augsburg

**Prof. Dr.-Ing. Prof. e. h. Wilhelm Bauer**, Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO und Universität Stuttgart

### **Mitglieder der Arbeitsgruppe sind:**

**Prof. Dr. Lars Adolph**, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)

**Prof. Dr.-Ing. Jan C. Aurich**, Technische Universität Kaiserslautern

**Vanessa Barth**, IG Metall

**Klaus Bauer**, TRUMPF Werkzeugmaschinen GmbH + Co. KG

**Nadine Bender**, KUKA Deutschland GmbH

**Prof. Dr. Angelika Bullinger-Hoffmann**, Technische Universität Chemnitz

**Prof. Dr.-Ing. Barbara Deml**, Karlsruher Institut für Technologie (KIT)

**Prof. Dr. Prof. h.c. Andreas Dengel**, Technische Universität Kaiserslautern und Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH

**Dr. Jan-Henning Fabian**, ABB AG

**Prof. Dr.-Ing. Sami Haddadin**, Munich School of Robotics and Machine Intelligence,  
 Technische Universität München  
**Prof. Dr. Michael Heister**, Bundesinstitut für Berufsbildung (BIBB)  
**Prof. Dr.-Ing. Rolf Hiersemann**, Hiersemann Prozessautomation GmbH  
**Dr. Norbert Huchler**, Institut für Sozialwissenschaftliche Forschung e. V. (ISF-München)  
**Dr. Nadine Müller**, Vereinte Dienstleistungsgewerkschaft (ver.di)  
**Dr. Rahild Neuburger**, Ludwig-Maximilians-Universität München  
**Dr.-Ing. Matthias Peissner**, Fraunhofer-Institut für Arbeitswirtschaft und Organisation (IAO)  
**Prof. Dr.-Ing. Annika Raatz**, Leibniz Universität Hannover  
**Prof. Dr.-Ing. Jürgen Roßmann**, Rheinisch-Westfälische Technische Hochschule Aachen  
**Prof. Dr. Christoph M. Schmidt**, RWI – Leibniz-Institut für Wirtschaftsforschung und  
 Ruhr-Universität Bochum  
**Dr. Anke Soemer**, Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.  
**Prof. Dr. Jochen Steil**, Technische Universität Braunschweig  
**Andrea Stich**, Infineon Technologies AG  
**Oliver Suchy**, Deutscher Gewerkschaftsbund (DGB)  
**Prof. Dr.-Ing. Sascha Stowasser**, Institut für angewandte Arbeitswissenschaft (ifaa)  
**Dr. Hans-Jörg Vögel**, BMW Group  
**Jochen Werne**, Prosegur Cash Services Germany GmbH

#### **Die Arbeitsgruppe wird unterstützt von:**

**Dr. Chi-Tai Dang**, Universität Augsburg  
**Dr.-Ing. Jan Harder**, Munich School of Robotics and Machine Intelligence,  
 Technische Universität München  
**Dr. Andreas Heindl**, Geschäftsstelle der Plattform Lernende Systeme  
**Dr.-Ing. Michael Wächter**, Technische Universität Chemnitz

#### **Redaktion**

**Dr. Andreas Heindl**, Geschäftsstelle der Plattform Lernende Systeme  
**Dr. Ursula Ohliger**, Geschäftsstelle der Plattform Lernende Systeme

## Über die Plattform Lernende Systeme

Lernende Systeme im Sinne der Gesellschaft zu gestalten – mit diesem Anspruch wurde die Plattform Lernende Systeme im Jahr 2017 vom Bundesministerium für Bildung und Forschung (BMBF) auf Anregung des Fachforums Autonome Systeme des Hightech-Forums und acatech – Deutsche Akademie der Technikwissenschaften initiiert. Die Plattform bündelt die vorhandene Expertise im Bereich Künstliche Intelligenz und unterstützt den weiteren Weg Deutschlands zu einem international führenden Technologieanbieter. Die rund 200 Mitglieder der Plattform sind in Arbeitsgruppen und einem Lenkungskreis organisiert. Sie zeigen den persönlichen, gesellschaftlichen und wirtschaftlichen Nutzen von Lernenden Systemen auf und benennen Herausforderungen und Gestaltungsoptionen.

# Literatur

---

- acatech (2016):** Neue autoMobilität – Automatisierter Straßenverkehr der Zukunft. Online unter: <https://www.acatech.de/publikation/neue-automobilitaet-automatisierter-strassenverkehr-der-zukunft/> (letzter Zugriff: 20.03.2020).
- acatech/Fachforum Autonome Systeme (2017):** Chancen und Risiken für Wirtschaft, Wissenschaft und Gesellschaft (Abschlussbericht). Online unter: <https://www.acatech.de/publikation/fachforum-autonome-systeme-chancen-und-risiken-fuer-wirtschaft-wissenschaft-und-gesellschaft-abschlussbericht> (letzter Zugriff: 20.03.2020).
- EU High-Level Expert Group (2019):** Ethics Guidelines for Trustworthy AI. Online unter: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> (letzter Zugriff: 20.03.2020).
- Huchler, Norbert (2019):** Assimilierte vs. Komplementäre Adaptivität. Grenzen teil-autonomer Systeme, in: Hirsch-Kreinsen, Hartmut/Karačić, Anemari (Hrsg.): Autonome Systeme und Arbeit. Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt, Bielefeld, S. 139–180.
- Huchler, Norbert (2016):** Die Grenzen der Digitalisierung. Neubestimmung der hybriden Handlungsträgerschaft zwischen Mensch und Technik und Implikationen für eine humane Technikgestaltung. In: „Digitalisierung, IT und Arbeit“ HMD Praxis der Wirtschaftsinformatik, Heft 53 (1), Wiesbaden: Springer, S. 109–123.
- Manzeschke, Arne/Weber, Karsten/Rother, Elisabeth/Fangerau, Heiner (2013):** Ethische Fragen im Bereich altersgerechter Assistenzsysteme. Ergebnisse der Studie. Online unter: <https://www.technik-zum-menschen-bringen.de/service/publikationen/ethische-fragen-im-bereich-altersgerechter-assistenzsysteme> (letzter Zugriff: 20.03.2020).
- Plattform Industrie 4.0 (2019a):** Industrie 4.0 gestalten. Souverän. Interoperabel. Nachhaltig. Fortschrittsbericht 2019. Online unter: <https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/hm-2019-fortschrittsbericht.html> (letzter Zugriff: 20.03.2020).
- Plattform Industrie 4.0 (2019b):** Technologieszenario Künstliche Intelligenz in der Industrie 4.0. Online unter: <https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/KI-industrie-40.html> (letzter Zugriff: 20.03.2020).
- Plattform Lernende Systeme (2019):** Künstliche Intelligenz und Diskriminierung: Herausforderungen und Lösungsansätze. Whitepaper der AG3 – IT-Sicherheit, Privacy, Recht und Ethik. Online unter: <https://www.plattform-lernende-systeme.de/publikationen-details/kuenstliche-intelligenz-und-diskriminierung-herausforderungen-und-loesungsansaeetze.html> (letzter Zugriff: 20.03.2020).

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung



## Impressum

### Herausgeber

Lernende Systeme –  
Die Plattform für Künstliche Intelligenz  
Geschäftsstelle | c/o acatech  
Karolinenplatz 4 | 80333 München  
[www.plattform-lernende-systeme.de](http://www.plattform-lernende-systeme.de)

### Gestaltung und Produktion

PRpetuum GmbH, München

### Stand

Juni 2020

### Bildnachweis

Westend61/gettyimages/Titel  
VICTOR/iStock/S. 9

Bei Fragen oder Anmerkungen zu dieser  
Publikation kontaktieren Sie bitte Johannes Winter  
(Leiter der Geschäftsstelle):  
[kontakt@plattform-lernende-systeme.de](mailto:kontakt@plattform-lernende-systeme.de)

Folgen Sie uns auf Twitter: @LernendeSysteme

### Empfohlene Zitierweise

Norbert Huchler et al. (Hrsg.): Kriterien für die menschengerechte Gestaltung der Mensch-Maschine-Interaktion bei Lernenden Systemen – Whitepaper aus der Plattform Lernende Systeme, München 2020.

Dieses Werk ist urheberrechtlich geschützt.  
Die dadurch begründeten Rechte, insbesondere die  
der Übersetzung, des Nachdrucks, der Entnahme von  
Abbildungen, der Wiedergabe auf fotomechanischem  
oder ähnlichem Wege und der Speicherung in Daten-  
verarbeitungsanlagen, bleiben – auch bei nur auszugs-  
weiser Verwendung – vorbehalten.