Normung und Standardisierung von KI-Systemen aus soziotechnischer Perspektive

Cecilia Colloseus

Der Normungsbedarf von KI-Systemen

Wenn sozialarbeiterische Arbeitsprozesse digital unterstützt und in ihrem Kontext ethisch wie rechtlich begründete Entscheidungen getroffen werden sollen, müssen Systeme, die auf künstlicher Intelligenz (KI) basieren, strengen Standards entsprechen. Gerade wenn es um Einschätzungen von Kindeswohlgefährdung und die Bewertung von Risikopotential geht, ist eine Ausrichtung an fundierten Normen unumgänglich.

Dieser Artikel befasst sich mit der Normung von KI. Normung ist ein wichtiger Schritt, um eine verantwortungsvolle und ethische Anwendung von KI-Technologien zu gewährleisten. Allerdings kann die digitale Operationalisierung in der Normung auch ethische Herausforderungen mit sich bringen, wie z.B. eine ungleiche Beteiligung von Interessengruppen oder die Verstärkung von bestehenden Vorurteilen und Diskriminierungen.

Die genannten Herausforderungen werden in diesem Artikel vor dem Hintergrund des EU-AI-Acts (AI-Act) und der DIN-KI-Normungsroadmap (DIN-KI) betrachtet. Es wird der Frage nachgegangen, ob die Vorgaben der Standardisierung und Normung eine Basis für die Umsetzung der in KAIMo verhandelten Fragestellung, ob ein Algorithmus moralisch kalkulieren kann, bieten können.

KI-Normung auf europäischer und nationaler Ebene – EU-AI-Act und DIN-KI-Normungsroadmap

Der EU-AI-Act ist ein 2021 von der Europäischen Kommission vorgestellter Vorschlag für eine neue Gesetzgebung zur Regulierung von KI. Der Gesetzesentwurf unterteilt KI-Systeme in drei Risikokategorien: Anwendungen, die ein nichtakzeptables Risiko darstellen (und folglich nicht in Umlauf gebracht werden dürfen), Hochrisikoanwendungen mit Regulierungsbedarf und Anwendungen mit geringem Risiko, die keiner weiteren Regulierung bedürfen. Der AI-Act sieht unter anderem die Einführung von verpflichtenden Anforderungen für Hochrisiko-KI-Anwendungen vor sowie die Verpflichtung für Anbietende von KI-Systemen, eine Risikoanalyse durchzuführen und eine umfassende Dokumentation über ihr KI-System bereitzustellen.

Die DIN-KI-Normungsroadmap hingegen ist eine Initiative des Deutschen Instituts für Normung (DIN) zur Entwicklung von Normen und Standards für Künstliche Intelligenz. Die Roadmap wurde 2020 erstmals veröffentlicht und definiert die wichtigsten Handlungsfelder für die Normung von KI-Technologien in Deutschland. Hierzu gehören unter anderem Themen wie Ethik und Sicherheit von KI-Systemen, Datenqualität und -zugang sowie Bildung und Qualifizierung im Bereich Künstlicher Intelligenz.

Beide Initiativen, der EU-AI-Act und die DIN-KI-Normungsroadmap, sind wichtige Schritte zur Entwicklung einer verantwortungsvollen und moralischen Anwendung von KI-Technologien und zeigen das wachsende Bewusstsein für die Bedeutung von KI-Normen und -Standards auf nationaler und internationaler Ebene. Im Kontext von Algorithmen, die im Konfliktfall moralisch kalkulieren sollen, muss ein Framework für eine ethische digitale Operationalisierung in der KI-Normung herausgearbeitet werden, das sowohl theoretisch fundiert als auch praktisch anwendbar ist. Dieses Framework soll dazu beitragen, dass die Normung von KI-Technologien nicht nur technisch korrekt, sondern auch ethisch verantwortungsvoll und gerecht erfolgt.

Es gibt bereits einzelne internationale Standards zu KI: ISO/IEC 22989:2022 (Informationstechnik – Künstliche Intelligenz), ISO/IEC DIS 23894:2022 (Informationstechnik – Künstliche Intelligenz – Risikomanagement), ISO/IEC 23053:2022 (Framework für Systeme der Künstlichen Intelligenz (KI) basierend auf maschinellem Lernen (ML), ISO/IEC DIS 42001 (Information Technology – Artificial intelligence – Management system), ISO/IEC 5259 (Artificial intelligence – Data quality for analytics and machine learning). Für die Implementierung von KI-Systemen in unterschiedlichen betrieblichen

Kontexten muss jedoch auch auf andere Normen und Standards zurückgegriffen werden, wie etwa DIN-Normen zu Ergonomie (z.B. DIN EN ISO 9241–112:2017, DIN EN ISO 9241–110:2020).

Im Folgenden werden Inhalte der DIN-KI-Normungsroadmap und des EU-AI-Acts referiert, die an diesen Anspruch anknüpfen und ein entsprechendes Framework bereitstellen können. Der Fokus liegt hierbei auf Kapitel 4.4 der DIN-KI-Normungsroadmap »Soziotechnische Systeme«. Es sei vorangestellt, dass es in der Normungsroadmap vor allem um die Implementierung von KI-Systemen in Unternehmen geht. Auf Institutionen und Behörden wird nicht gesondert eingegangen und es ist zu prüfen, ob sich die entsprechenden Normen auch auf das im Rahmen von KAIMo untersuchte Feld anwenden lassen.

Soziotechnische KI-Normung

> »Vom Produktentstehungsprozess über Inbetriebnahme und alltäglicher operativer Anwendung bis hin zur Außerbetriebsetzung sind nicht nur der Stand der technologischen Entwicklung sowie der spezifische Anwendungsfall zu berücksichtigen, sondern auch die Grundsätze und Prinzipien einer menschengerechten und partizipativen soziotechnischen Gestaltung. Dieses Erfordernis spiegelt sich bislang meist nicht in den korrespondierenden Normen wider.« (DIN-KI, 162)

Da es im hier verhandelten Kontext um den Gebrauch von Algorithmen in sozialen Konfliktfällen geht, werden die technischen Aspekte der KI-Normung ausgespart und dafür die soziotechnischen Zusammenhänge prominent hervorgehoben. Als soziotechnische Systeme werden solche Systeme bezeichnet, in denen Mensch und Technik miteinander verknüpft sind und in Wechselwirkung zueinander stehen (Zweig et al. 2021, Schlick et al. 2010, Suchman 2009). Auch KI-Technologien sind in diesem Kontext zu betrachten. Der einzelne Mensch, sein organisatorisches Umfeld und die Gesellschaft als Ganzes müssen in den Blick genommen werden, wenn Regeln und Normen hierfür aufgestellt werden sollen.

Im EU AI-Act wird gefordert, dass Hochrisikosysteme mit Funktionen ausgestattet werden, die den Menschen aktiv einbinden. So soll es für alle Betroffenen und Beteiligten ein hohes Maß an Transparenz geben, die menschliche Aufsicht soll gewährleistet und eine Art »Stopptaste« eingerich-

tet werden, die vom Menschen ausgelöst werden kann (EU AI-Act, Artikel 14, 4d). Der Mensch steht also immer im Mittelpunkt. Alle technischen Komponenten müssen sich an den soziotechnischen Anforderungen ausrichten und an der gesamten Entwicklung müssen alle relevanten Akteur*innen beteiligt werden (partizipative Forschung und Design).

In ihrer zweiten Ausgabe adressiert die DIN-KI-Normungsroadmap soziotechnische Aspekte wie die Mensch-Technik-Interaktion, die humane Arbeitsgestaltung sowie Anforderungen an Unternehmensstrukturen und -prozesse in einem eigenen Kapitel (DIN-KI, 153–175). Bereits in den allgemeinen Handlungsempfehlungen zu Beginn wird die Empfehlung ausgesprochen, »[d]en Menschen als Teil des Systems [zu] begreifen, und zwar in allen Phasen des KI-Lebenszyklus.« (DIN-KI, 35). Die Normungsroadmap fordert außerdem, »[g]esellschaftliche und ethische Fragestellungen mithilfe etablierter Modelle […] zu operationalisieren, messbar bereits bei der Entwicklung der Technologie zu verankern und dabei auf dem Stand der Forschung zu Diskriminierungssensibilität aufzubauen« (DIN-KI, 36). Des Weiteren ist vorgesehen, dass eine »adäquate Organisationskultur« etabliert wird. Konkret bedeutet das, dass in den individuellen Unternehmen oder Institutionen die relevanten Akteur*innen sensibilisiert, qualifiziert und in einem geeigneten Change Management im Prozess mitgenommen werden (DIN-KI, 36).

Ein weiterer Anspruch, den die Roadmap in diesem Kontext formuliert, ist, dass die beteiligten Menschen über den gesamten Lebenszyklus von KI-Systemen hinweg mit Prozessen, Methoden und Tools unterstützt werden sollen. Das setzt voraus, dass die vom System betroffenen und daran beteiligten Menschen auch in alle Phasen des KI-Lebenszyklus eingebunden werden. Normungsgremien müssen dafür konkrete Normen erarbeiten, die gewährleisten, dass die notwendigen Transparenzanforderungen und vor allem die menschliche Aufsicht über KI-Systeme eingehalten werden. Auch in der Erarbeitung dieser Normen müssen Menschen aus den unterschiedlichen betroffenen Zielgruppen berücksichtigt werden (z.B. Zivilgesellschaft).

Die soziotechnische Perspektive zeigt auf, dass es bei KI nicht allein um die technische Machbarkeit geht, sondern dass der Kontext der jeweiligen Anwendung beachtet werden muss. Sie ist das »multiperspektivische ›Gegengewicht‹ zu einer rein technikzentrierten Sicht auf KI« (DIN-KI, 156).

Entwicklung von KI-Systemen nach soziotechnischen Kriterien

Laut der KI-Normungsroadmap reicht es nicht aus, die soziotechnische Perspektive nur in der Entwicklung von KI-Systemen einzunehmen. Sie muss vielmehr über den gesamten Lebenszyklus des Systems betrachtet werden (vgl. ISO/IEC 22989:2022, ISO/IEC 23053:2022), da sich KI-Systeme in einem gewissen Rahmen weiterentwickeln. Dieser Rahmen muss in der Designphase abgesteckt werden. Ist das System bereits in Betrieb, kann immer nur ein Ausschnitt des Systemzustands zu einem bestimmten Zeitpunkt abgebildet werden.

In der ersten Phase des KI-Systems, der Initialisierung, werden Ziel und Zweck der Anwendung definiert (vgl. ISO/IEC 22989:2022). Es wird geklärt, welche Anforderungen die KI innerhalb des soziotechnischen Systems erfüllen muss, also welches Problem sie lösen soll, welche Bedürfnisse ihre Zielgruppe hat und welche Erfolgsparameter es gibt. Dabei geht es nicht (nur) um die technische Machbarkeit, sondern »darum, auf Basis einer eingehenden Problemanalyse in einer gegebenen Situation von der Idee zur Entscheidung für ein KI-System zu gelangen und den Entwicklungsprozess anzustoßen« (DIN-KI, 159). Hier müssen auch ethische Aspekte einbezogen werden, z.B. anhand von ethics-by-design-Katalogen wie dem der Bertelsmann Stiftung (Puntschuh und Fetic 2020). Auch das WKIO-Modell der AI Ethics Impact Group liefert eine Methode, vorab definierte ethische Werte zu operationalisieren (VDE, Bertelsmann Stiftung 2020). Teil dieser Initialisierungsphase ist außerdem eine erste Risikoanalyse, die die soziotechnischen Folgen aus Sicht mehrerer Stakeholder identifiziert (ISO/IEC 23894:2022). Hier geht es nicht nur um technische und rechtliche Fragen, sondern auch und vor allem um ethische und soziale Folgen. So müssen die Fragen geklärt werden, welche Grundrechte oder -werte durch den Einsatz der Software potenziell berührt werden, was die beabsichtigten Auswirkungen der Software sind, wer vom Einsatz des algorithmischen Assistenzsystems betroffen ist, welche potenziellen Auswirkungen der Einsatz der Software auf die unterschiedlichen Stakeholder, Gesellschaft, Wirtschaft oder Umwelt hat, welche Risiken durch mögliche Fehler bei der Entwicklung oder dem Einsatz der Software entstehen können und welche Szenarien hier denkbar sind (Puntschuh und Fetic 2020).

Participation is Key

Bevor die Entscheidung für ein KI-System getroffen wird, muss definiert werden, welche Personen (-gruppen) überhaupt davon betroffen sein werden und welche Bedürfnisse sie haben. Diese Personen müssen an der Entwicklung und Implementierung des Systems unbedingt beteiligt werden (DIN-KI, 160). Nur, wer die Zielgruppe kennt, kann ein System entwickeln, das divers ist und keine stereotypischen Vorstellungen von Menschen (-gruppen) reproduziert. Ein möglicher Ansatz für die Einbindung von unterschiedlichen Betroffenen ist der Participatory-Design-Ansatz (vgl. Simonsen und Robertson 2013). Gemeinsam werden Szenarien erarbeitet, die sich im jeweiligen Kontext mit Implementierung des KI-Systems ergeben können. So werden z.B. gemeinsam Prototypen entwickelt oder bestimmte Situationen simuliert. Nutzende erhalten auf diese Art ein besonderes Mitspracherecht, ohne selbst entsprechende technische Fachkenntnisse haben zu müssen. Im folgenden Entwicklungsprozess werden sie an der Evaluation des Systems beteiligt. Hier bieten sich Ansätze von XAI (Explainable AI) an, die über Visualisierung einen Zugang zu den zugrunde liegenden Softwarelogiken ermöglichen oder kritische Entscheidungspunkte offenlegen. Grundsätzlich wäre die idealtypische Forderung bei der Planung und Gestaltung einer KI-Lösung, eine Vertretung aller Stakeholder zu beteiligen.

Die ISO/IEC 22989:2022 unterteilt Stakeholder in: »AI-Provider«, »AI-Producer«, »AI-Customer«, »AI-Partner*innen«, »AI-Subject« und »Relevant Authorities«. Einzubeziehen sind z.B. Expert*innen mit Domänenwissen (KI-Expert*innen, Data Scientists, Informatiker*innen usw., Prozessgestaltende, Usability- Expert*innen, Produktgestaltende usw., Softwaretester*innen, Ergonom*innen, Psycholog*innen usw., Sicherheitsexpert*innen in der jeweiligen Domäne, Expert*innen für Ethik, Diversity, Fairness usw.), Expert*innen aus den betroffenen Fachabteilungen, Nutzende des KI-Systems, Interessensvertretungen von Betreibenden und Nutzenden, Entscheider*innen über den Einsatz der KI-Lösung, Vertreter*innen der Zivilgesellschaft, aber auch »überraschende« Stakeholder, also solche Personen, die nicht unmittelbar mit dem jeweiligen System zu tun haben, aber mittelbar betroffen sein können. Wie und in welchem Kontext die jeweiligen Stakeholder eingebunden werden, ist abhängig vom Zeitpunkt im Projektlebenszyklus, also während der Zielfindung, während der Planung und Gestaltung, während der Inbetriebnahme, im laufenden Betrieb bzw. im kontinuierlichen Verbesserungsprozess.

Die Kommunikation muss dabei immer zielgruppengerecht und inklusiv sein. So ist die Kommunikation zwischen den Expert*innen mit Domänenwissen untereinander, zwischen den Expert*innen mit Domänenwissen und den Nutzenden, zwischen Nutzenden und Technik, zwischen Expert*innen und sonstigen Beteiligten jeweils unterschiedlich. Um die Beteiligung gut durchzuführen, können einschlägige Normen und Standards (z.B. VDI-MT 7001:2021) herangezogen werden. Zu beachten ist darüber hinaus, dass die Beschäftigten entsprechende Kompetenzen erwerben müssen, um mit dem System richtig umgehen zu können. Entsprechende Schulungen müssen frühzeitig durchgeführt und am Wissensstand der Beschäftigten ausgerichtet werden.

KI-Systeme – Werkzeuge oder Agenten?

Wenn IT-Systeme (insbesondere KI) soziotechnisch gestaltet werden, müssen sie geeignet sein, (Arbeits-)Aufgaben von Menschen in unterschiedlichen Rollen und im Nutzungskontext zu unterstützen. Das bedeutet zum Beispiel, dass die Schnittstellen ergonomisch gestaltet werden müssen. Die Nutzenden sollen vom KI-System dabei unterstützt werden, Aufgaben effektiv, effizient und zufriedenstellend zu erledigen. Dabei müssen zunächst die Bedürfnisse der Nutzenden erkannt und analysiert werden. Methoden der partizipativen Sozialforschung sind hier das Mittel der Wahl.

In der Normungsroadmap wird konstatiert, dass KI-Systeme nicht nur als technische Werkzeuge wahrgenommen werden. Vielmehr sollen sie als »eine neue Klasse von Agenten in der Organisation« (Raisch und Krakowski 2020) betrachtet werden. Die Interaktion zwischen Menschen und diesen neuen nichtmenschlichen Agenten wird in Autonomiestufen unterteilt. So soll die KI etwa nur dann eine Aufgabe erledigen können, wenn ein Mensch die Bestätigung dafür gibt, oder es muss möglich sein, dass der kontrollierende Mensch ein Veto gegen die Entscheidungen einer KI einlegt. Auf einer höheren Autonomiestufe handelt die KI vollständig autonom und der Mensch wird nur dann informiert, wenn er konkret nachfragt. Die höchste Stufe von Autonomie ist dann erreicht, wenn die KI handeln kann, ohne den Menschen einzubeziehen.

Bislang wurden in Konzepten wie Ergonomics/Human Factors (EHF) statische technische Systeme betrachtet (z.B. stationäre Maschinen). KI-Systeme sind jedoch inhaltlich und zeitlich dynamisch. Deshalb muss das EHF-Gestal-

tungskonzept erweitert werden, damit die Dynamik von Schnittstellen, Funktionsweisen und Auswirkungen auch für Menschen passend gestaltet werden können. So werden etwa auch menschliche Eigenschaften wie Empathie mit einbezogen (André und Bauer 2021, Höddinghaus et al. 2021, Brynjolfsson et al. 2018, Moray 1989). Werden Systeme nach soziotechnischen Grundsätzen gestaltet, werden Technologieeinsatz und Organisation gemeinsam optimiert (»joint optimization«) (vgl. Cherns 1976&1987, Ulich 2013).

Die KI als gesellschaftliche Akteurin

Neben der Organisation und dem (individuellen) Menschen, spielt in soziotechnischen Systemen auch die umgebende Gesellschaft eine entscheidende Rolle. Sie bildet die Schnittstelle von Individuum und Organisation. Bestehende Ungleichheitsverhältnisse oder Diskriminierung, die innerhalb der Gesellschaft bestehen, können sich in KI-Anwendungen verfestigen. Technologien, die menschliche Intelligenz nachahmen sollen, können deshalb niemals objektiv oder neutral sein (Benjamin 2019), da sie eben immer in eine von bestimmten Werten – und damit einhergehend: sozialen Herausforderungen – geprägten Gesellschaft eingebettet sind. Nutzen Menschen solche Technologien, beeinflussen sie diese und umgekehrt (Suchman 2007). Besonders deutlich wird das beim Maschinellen Lernen: Hier reagieren Softwareprogramme dynamisch und adaptiv auf ihre Nutzer*innen (Pentenrieder et al. 2020). Betrachtet man Mensch und Maschine auf diese Weise, wird die bisherige Konzeption von autonomen und strikt trennbaren Entitäten in Frage gestellt. Menschliche und maschinelle Akteur*innen ergeben durch Kollaboration und Interaktion ein Ganzes (Suchman 2007).

Training des KI-Systems: Das Datenset

Wird ein konsistentes soziotechnisches Mensch-Technik-Organisation-Modell verwendet, müssen zentrale Fragen der Einführung, Nutzung und Folgeabschätzung von KI bearbeitet werden (Huchler et al. 2020, Stowasser und Suchy 2020). Die erste Frage betrifft dabei die Daten, auf deren Grundlage die KI entwickelt wird und den Zweck, dem sie dienen soll. Als zweites stellt sich die Frage, wie das menschliche Verhalten durch den Einsatz der KI beeinflusst wird z.B. Autonomie oder Entscheidungsdilemmata. Die dritte Frage betrifft

das Verhältnis zwischen KI und menschlichen Bedürfnissen und die vierte Frage die systemischen Folgewirkungen. Hier geht es sowohl um die Wirkungen innerhalb des Systems als auch in den Subsystemen, der Systemumwelt und Gesellschaft (DIN-KI, 155). Zur ersten Frage sei erläutert: KI-Systeme müssen anhand spezifischer Daten trainiert werden, um in den betrieblichen Einsatz überführt werden zu können. Die Auswahl der Trainings-, Validierungs- und Testdaten muss dabei so erfolgen, dass Diskriminierung vermieden wird. Außerdem müssen die Daten auf ihre Qualität hinsichtlich des geplanten Einsatzes überprüft werden: Sind genug Daten vorhanden, sind sie konsistent, sind die Daten aktuell, befinden sich falsche Daten im Set etc.? Auswahl, Training, Verifizierung und Validierung der verwendeten Datensätze und die Testung der KI-Lösung sind adäquat zu dokumentieren.

Es ist hinlänglich bekannt, dass Menschen systematisch Entscheidungsfehler machen (Kahnemann und Tversky 1982). In dieser Hinsicht sind ihnen KI-Systeme ähnlich, denn auch sie können in ihrer Entwicklung und im Einsatz Bias-Effekte oder Entscheidungsfehler bezüglich der Fairness verursachen. Als »Bias« werden unerwünschte Verzerrungen bezeichnet, die entweder bereits bei der Erhebung der Datensätze oder durch die Selektion bzw. Art ihrer Verarbeitung aufkommen können. Diese Bias-Effekte sind jedoch bedingt durch menschliche Designentscheidungen (z.B. Datenbank und Logik) und die dem jeweiligen Einsatzgebiet zugrunde liegenden Vorannahmen. Die Rechenschaftspflicht und die Gewährleistung von Fairness liegen also beim Menschen. Bei Bias-Effekten sind die Faktoren Unsicherheit und Risiko zu beachten. Als Risiko wird bezeichnet, wenn ein bestimmter bekannter Umweltzustand mit einer empirisch ermittelten Wahrscheinlichkeit eintreffen kann. Risiko ist also quantifizierbar und potentiell steuerbar. Unsicherheit hingegen bedeutet, dass weder die möglichen Umweltzustände noch die mögliche Eintrittswahrscheinlichkeit bekannt sind (DIN-KI, 156). Algorithmen für Risikosituationen und -abschätzung wägen das Risiko in Form von Eintrittswahrscheinlichkeiten und gewünschten Optimierungslevels ab (Kahnemann und Tversky 1982). Dabei wird deutlich, dass die menschliche Wahrnehmung bei der Analyse, Gestaltung und Bewertung von KI-Systemen eine entscheidende Bedeutung hat. Zur Einschätzung der Risiken gehört auch, Ansprüche an Transparency und Rechenschaftspflicht zu formulieren. Welche Informationen müssen für wen offengelegt werden? Mit welcher technischen Tiefe müssen Informationen angereichert werden, um gleichzeitig hilfreich und verständlich zu sein? Wer ist rechenschaftspflichtig im Schadensfall? Im Anschluss an das Risiko-Assessment müssen Kosten, Aufwand

und Ressourcen für die Umsetzbarkeit der Anwendung ermittelt werden (vgl. ISO/IEC 22989:2022).

Soziale Nachhaltigkeit im soziotechnischen Kontext

KI-Anwendungen müssen bestimmten Nachhaltigkeitskriterien genügen und parametrisierbar sein in Bezug auf quantitative Zielsetzungen.

> »Im Hinblick auf die Entwicklung, Nutzung und den Einsatz von KI-Systemen bedeutet Nachhaltigkeit vor allem, dass die Würde des Menschen respektiert wird, keine Menschen ausgeschlossen, benachteiligt oder diskriminiert werden und die menschliche Autonomie und Handlungsfreiheit durch KI-Systeme nicht eingeschränkt werden dürfen. In einer erweiterten Perspektive auf Nachhaltigkeit bedeutet soziale Nachhaltigkeit auch, dass neben körperlicher Unversehrtheit und menschenwürdigen Lebensbedingungen auch die Fähigkeit, auf menschliche Art und Weise zu denken, zu argumentieren und zu handeln, nicht eingeschränkt werden sollte. Hier zeigt sich schon, dass ein umfassendes Verständnis von sozialer Nachhaltigkeit sehr weitreichende Konsequenzen für die Gestaltung von KI-Systemen hat.« (Rohde et al. 2021).

Gesetze und Normen können nicht jedes Detail regeln. Subsidiäre Aushandlungssysteme, z.B. auf betrieblicher Ebene und individuelle Entscheidungsrechte müssen hier ergänzend eingesetzt werden. Muhammad (2022) konstatiert, dass KI-Systeme Individuen und Gruppen Schaden zufügen können und etabliert verschiedene Typen von Fehlern, die in diesem Kontext passieren können:

- »Vergabe-Fehler«: Das System hält Möglichkeiten, Ressourcen oder Informationen zurück oder stellt sie unfair zur Verfügung.
- »Servicegüte-Fehler«: Das System arbeitet nicht für alle Gruppen ähnlich gut.
- »Repräsentations-Fehler«: Die Entwicklung oder die Verwendung eines Systems repräsentiert Gruppen unterschiedlich.
- »Stereotyp-Fehler«: Das System reproduziert und verstärkt Stereotypen, indem beispielsweise stereotypische Charakteristika unreflektiert allen Angehörigen einer Gruppe zugewiesen werden.

- »Verunglimpfungs-Fehler«: Das System wird aktiv abwertend oder beleidigend.
- »Prozess-Fehler«: Das System trifft Entscheidungen aufgrund von Charakteristika, die nicht für die Aufgabe relevant sein sollten (DIN-KI, 156f.)

Im Datenschutz müssen die Grundprinzipien Zweckfeststellung bzw. Zweckbindung von Daten, Erforderlichkeit der Datensammlung, Transparenz, Datenvermeidung und Datensparsamkeit eingehalten werden. Diese Prinzipien gelten auch für die Gestaltung soziotechnischer Systeme.

Für ethische Betrachtungen wurden Gütesiegel vorgeschlagen, die auf Wertanalyseverfahren aus einer Kombination von Zielkriterien, Indikatoren und messbaren Größen beruhen. Die Bewertung dieser normativen Anforderungen sollten sich auf technische Prüfungen stützen können.

Anwendungsspezifische Anforderungen bringen normative Grundsätze in die konkrete Anwendung und fügen spezielle Einsatzanforderungen hinzu. Sie bilden die Grundlage der Risikoeinstufung gemäß AI Act, greifen dabei relevante ethische Aspekte auf, nutzen das New Legislative Framework, um komponentenweise die Konformitätsvermutung von Herstellenden zu unterstützen und formulieren im Grundsatz Anforderungen an das gesamte technische System, in das KI eingebettet ist. Hierbei geht es um die vertikale Bewertung von KI, bei der geprüft wird, ob die KI für einen bestimmten Einsatzzweck geeignet ist (DIN-KI, 128).

Funktionsteilung Mensch-Technik

In der Funktionsteilung zwischen Mensch und KI-System gilt das Primat der (Arbeits)Aufgabe. Die Gestaltung der Aufgabe steht also am Anfang des Gestaltungsprozesses und ordnet ihr die Gestaltung der Ausführungsbedingungen unter (Hacker und Sachse 2014, Ulich 2011, DIN EN 614–2:2008). Der Autonomiegrad des KI-Systems wird repräsentiert durch die Funktionsteilung. Die einschlägigen Normen sind dahingehend zu prüfen, ob sie die verschiedenen Autonomiegrade berücksichtigen. Für die Funktionsteilung wird die von Fitts (1951) entwickelte HABA MABA-Liste (= »humans are better at« – »machines are better at«) herangezogen. Es kann aber keine starre Funktionszuteilung zwischen Mensch und Technik vorgenommen werden, da diese Subsysteme nicht mechanistisch zusammenwirken. Außerdem pauschaliert eine solche verkürzende Dartstellungsweise Fertigkeiten, Fähigkeiten und Wissen

des Subsystems Mensch und beachtet die Dynamik und Weiterentwicklung der Subsysteme nicht. Auch wird die Lebenszyklusperspektive für beide Subsysteme nicht berücksichtigt. Die Funktionsteilung kann sich aber abhängig von der Situation dynamisch ändern, z.B. wenn der Mensch in einer Gefahrensituation eingreifen muss. Zu dieser Adaptivität gibt es aktuell noch keine Normung.

In der dynamischen Funktionsallokation kann es zu »Ironien der Automatisierung« kommen (vgl. Bainbridge 1987). Das bedeutet, dass mit der Automatisierung die Systemkomplexität steigt und neue Aufgaben der Überwachung, Steuerung und Korrektur entstehen, für die der Mensch nicht ausreichend qualifiziert ist. Wenn Funktionsteilung und Automatisierung gestaltet werden, muss diese »Ironie« berücksichtigt werden und in die relevanten Normen und Standards einfließen.

Entscheidend ist außerdem das Menschenbild:

> »Wird der Mensch vom Entwickelnden als Fehlerquelle gesehen, so wird die Gestaltung tendenziell versuchen, den Einfluss des Menschen im KI-System weitgehend zu reduzieren. Wird das KI-System hingegen als Unterstützung für den Menschen betrachtet, so wird die Funktionsteilung eher komplementär erfolgen.« (DIN 166)

Im Interesse einer menschenzentrierten KI-Nutzung sollte dem Leitbild einer komplementären Funktionsteilung der Vorzug gegeben werden.

Um ein KI-System sinnvoll nutzen zu können, müssen auch die Organisation und Prozesse, in die es eingebettet ist, auf eine bestimmte Weise gestaltet werden: So muss ein Vertrauen in die Automation etabliert werden und die Potentiale für Belastung und Beanspruchung durch die Nutzung des KI-Systems (z.B. Technikstress) erhoben werden (vgl. DIN-EN-ISO-10075-Reihe). Es müssen systemische Effekte mitbedacht werden (z.B. Aufschaukelungseffekte durch den Eingriff des Menschen in das KI-System) und die veränderte Risikokompensation der Nutzenden sowie deren Folgen beim (unbemerkten) Ausfall des Systems.

Die Qualifizierung der Nutzenden, die partizipative Gestaltung und ein geeignetes Change Management sind hier entscheidend. Es ist zu prüfen, inwieweit diese Aspekte in den relevanten Normen und Standards (z.B. DIN EN ISO 27500:2017, VDI/VDE-MT 7100, DIN EN ISO 9001:2015) abgebildet sind.

Grundsätzlich sind bezogen auf Mensch-Technik-Interaktionen in soziotechnischen Systemen drei hierarchisch strukturierte Schnittstellen mit

jeweils darauf bezogenen Gestaltungsprinzipien von besonderem Interesse: Aufgabenschnittstelle, z.B. nach DIN EN 614–2:2008, Reihe DIN EN ISO 11064:2011, Interaktionsschnittstelle, z.B. nach DIN EN 894–1:2009, DIN EN ISO 9241–11:2018, DIN EN ISO 9241- 110:2020, ISO/IEC 29138–1:2018, Informationsschnittstelle, z.B. nach VDI/VDE 3850–1:2014, ISO 9241–112:2017).

Implementierung der KI-Lösung im soziotechnischen System

Ist die Entscheidung für den KI-Einsatz gefallen und die Grob- und Feinplanung der KI-Lösung abgeschlossen, muss ihre Inbetriebnahme anhand von Projektmanagement-Normen geplant werden (z.B. DIN ISO 21500:2016, Reihe DIN 69901, Reihe DIN 69909). Dabei muss geprüft werden, ob KI-Projekte hinsichtlich des Projektmanagements Besonderheiten aufweisen, die bei Bedarf in der Normung abzubilden sind.

Die weiteren Entwicklungsschritte erfolgen iterativ. So können Informationen, die im Betrieb des Systems neu hinzukommen, eine Rückkehr zu den Schritten in der Initialisierungsphase erforderlich machen. Der Prozess orientiert sich an den Schritten »Design und Entwicklung« und »Verifikation und Validierung« der ISO/IEC 22989:2022. Die in dieser Phase entwickelte Groblösung des KI-Systems wird für die Inbetriebnahme vorbereitet. Alle Beteiligten, etwa die Betreibenden des KI-Systems, spätere Nutzende des KI-Systems, Interessensvertretungen von Betreibenden und Nutzenden, Vertretende der Zivilgesellschaft, aber auch »überraschende Stakeholder«, müssen in dieser Phase einbezogen werden. Nur so kann die Umsetzung von ergonomischen Grundsätzen und Prinzipien sowie eine gebrauchstaugliche Gestaltung von Produkten und Arbeitsmitteln gewährleistet werden.

Mensch, Technik, Organisation

Soziotechnische Systeme werden immer von sachlichen (technisch-organisatorischen) und menschlichen (persönlichen) Gegebenheiten beeinflusst (z.B. DIN EN ISO 6385:2016). Deshalb sind Mensch, Technik und Organisation stets in ihrer gegenseitigen Abhängigkeit und ihrem Zusammenwirken zu reflektieren (MTO-Konzept). Die Gestaltungsdimensionen eines KI-Systems ergeben sich daher also aus den Elementen Mensch, Technik und Organisation sowie aus deren Schnittstellen (Mensch-Technik, Mensch-Organisation und Organi-

sation-Technik). Aus den identifizierten Gestaltungsdimensionen resultieren verschiedene Normen und Standards, die für die Planung und Gestaltung des jeweiligen Systems heranzuziehen sind.

Bei der Planung und Gestaltung jedes KI-Systems ist zwingend eine Analyse des zugrunde liegenden soziotechnischen Systems durchzuführen. Im Arbeitskontext ist das soziotechnische System das »Arbeitssystem« (vgl. Ulich 2013). Laut DIN EN ISO 6385:2016 ist das Arbeitssystem ein »System, welches das Zusammenwirken eines einzelnen oder mehrerer Arbeitender/Benutzer*innen mit den Arbeitsmitteln umfasst, um die Funktion des Systems, innerhalb des Arbeitsraumes und der Arbeitsumgebung unter den durch die Arbeitsaufgaben vorgegebenen Bedingungen, zu erfüllen«.

Ethische Aspekte sind bei der Planung und Gestaltung des soziotechnischen Systems stets zu beachten und für den gesamten Lebenszyklus des KI-Systems zu gestalten. Dazu zählen z.B. Transparency, Accountability, Privacy, Justice, Reliability und Sustainability (z.B. AI Ethics Label der AI Ethics Impact Group, VDE 2020).

Die einschlägigen Normen bzw. Standards zu KI (z.B. ISO/IEC 22989, ISO/IEC 42001, DIN SPEC 92001 Reihe, DIN SPEC 92001–2:2020, DIN SPEC 92001–3, ISO IEC 25059:2022), Ergonomie & Organisation (z.B. DIN EN ISO 6385:2016, DIN EN ISO 26800:2011, DIN EN ISO 9241 Reihe, Reihe, DIN EN ISO 27500:2017, VDI/VDE-MT 7100) und Ethik (VDE SPEC 90012, IEEE 7000 Serie, ISO IEC/TR 24028) berücksichtigen die resultierenden Anforderungen aus der soziotechnischen Gestaltung eines KI-Systems noch nicht hinreichend und lassen oft die Wechselwirkungen zwischen Mensch, Technik und Organisation außer Acht.

Überprüfung des KI-Systems im Regelbetrieb

In regelmäßigen Abständen muss überprüft und entschieden werden, ob die gewünschte Funktionsweise an geänderte Rahmenbedingungen angepasst werden muss. Im Betrieb können Echtdaten (je nach Anwendungsfeld anonymisiert oder pseudonymisiert) gesammelt und für die relevanten Akteur*innen im soziotechnischen System verständlich und transparent aufbereitet werden. So kann das System kontinuierlich verbessert werden.

Auch in diese kontinuierliche Evaluierung und Anpassung des Systems sollten die Betroffenen eingebunden werden. Ihre Erfahrungen sind die Grundlage für nötige Verbesserungen (Stowasser und Suchy 2020). Da-

für wird eine technische Lösung mit einem Transparency-by-Design- bzw. Transparency-by-Default-Ansatzes benötigt, die Menschen ermöglichen, den Überblick zu behalten. Das kann als Modul im System erfolgen oder durch ein eigenständiges Command Tool.

Die Stopptaste

Personen, die innerhalb des Systems eine Expert*innenrolle einnehmen, werden unter dem Begriff High Level Expert Group (HLEG) zusammengefasst. Sie gewährleisten die menschliche Aufsicht über das KI-System und werden dann als »Human-in-the-Loop (HITL, im Entscheidungszyklus der KI involviert), Human-on-the-Loop (HOTL, beim Design der KI und im Monitoring involviert) oder »Human in Command« (HIC, soll die Gesamtaktivität inklusive breitere ökonomische, soziale, rechtliche und ethische Auswirkungen überblicken können)« (DIN-KI, 169) bezeichnet. Der HIC wird im EU-AI-Act prominent vorgestellt und soll v.a. für Hochrisikosysteme, über geeignete Interventionsmöglichkeiten verfügen. Dazu zählt auch das Auslösen einer »Stopptaste« für die KI (Heesen et al. 2021). »Stopptaste« bedeutet nicht, eine KI-Prozedur bei Zweifeln zu unterbrechen, sondern »die Möglichkeit, einer durch KI getroffenen Entscheidung nicht zu folgen oder die KI-Nutzung für einen bestimmten Zeitraum auszusetzen und stattdessen Menschen entscheiden zu lassen«. (DIN-KI, 169)

Menschliche Interventionen sollten immer möglich sein. Etwa sollten Menschen Ausnahmen von den Entscheidungen der KI treffen können, oder Parameter des Systems (Schwellenwerte, Eingangsgrößen) rekonfigurieren. Der Ansatz »keep the human in the loop« betrachtet einzelne Individuen im Verhältnis zur KI. Demgegenüber steht das Gestaltungsprinzip »keep the organization in the loop«. Es sollen also auch die Interaktion der relevanten Stakeholder betrachtet und laufend optimiert werden (Hermann 2020).

Erklärbarkeit

Es ist von großer Bedeutung, dass KI-Systeme erklärt werden können (Explainable AI, XAI). Nutzende sollten verstehen können, welche Inputs zu welchen Outputs führen, welche Aspekte im Vorhinein festgelegt werden können und welche durch Erklärbarkeitsmetriken identifiziert oder im Nachhinein

über Zielkorridore ermittelt werden. Bei der Entwicklung von soziotechnischen Systemen werden Ziele und Maßnahmen unter Berücksichtigung von Folgeabschätzungen definiert, aber nicht alle Entscheidungen können im Voraus getroffen werden, da sich Rahmenbedingungen ändern können und unbeabsichtigte Effekte auftreten können. Veränderungen des Datensatzes im Betrieb des KI-Systems können durch Methoden im Bereich »Drift« erkannt werden und sollten eine Standardfunktion im KI-System sein. Eine kontinuierliche Überprüfung und Bewertung der Ziele und Entscheidungen in Bezug auf das KI-System ist notwendig und sogar verpflichtend für Hochrisikosysteme gemäß AI Act, Art. 61. Dabei müssen wichtige Fragen beantwortet werden, wie z.B. ob weitere Ziele berücksichtigt werden müssen, ob die Funktionsweise noch korrekt ist und ob Probleme erkannt werden können. »Near Misses« (beinahe Ausfälle) sind oft wertvolle Einblicke in das System, wenn es an seinen Grenzen betrieben wurde. Es ist wichtig, Berichtsstrukturen über Ausfälle oder fast-Ausfälle zu schaffen, um KI-Systeme zu verbessern und zukünftige Systemresilienz sicherzustellen oder Risiken zu simulieren. Es ist auch ratsam, regelmäßig zu überprüfen, ob es grundlegende Änderungen in der KI-Technologie gibt, die die eigene Lösung verbessern oder ersetzen könnten.

Zusammenfassung und Ausblick

Wie der EU-AI-Act und die DIN-Normungsroadmap KI zeigen, ist die Notwendigkeit von Normung und Standardisierung im Kontext von KI von den zuständigen Gremien erkannt worden. Wie bei allen technischen Entwicklungen zuvor, wird es auch für KI-Anwendungen unvermeidbar sein, Zertifizierungen und entsprechende Audits durchzuführen. KI-Systeme, die dabei helfen sollen, im Konfliktfall moralisch zu kalkulieren, fallen in die Kategorie »Hochrisikosystem« und müssen strengen Normen und Standards entsprechen. Diese müssen mit Bedacht festgesetzt werden. Wie in diesem Beitrag dargelegt wurde, ist es entscheidend, vor Einführung eines KI-Systems die tatsächlichen Bedarfe zu identifizieren. Allen voran muss die Frage geklärt werden, ob der Einsatz einer KI wirklich notwendig und sinnvoll ist, oder ob eine andere Lösung gefunden werden kann. Fällt die Entscheidung zugunsten der KI aus, müssen die davon Betroffenen in jeden Schritt der Entwicklung der individuellen KI-Lösung einbezogen werden.

Literatur

André, Elisabeth, and Wilhelm Bauer. 2021. *Kompetenzentwicklung für Künstliche Intelligenz – Veränderungen, Bedarfe und Handlungsoptionen*. Whitepaper aus der Plattform Lernende Systeme, München. DOI: https://doi.org/10.48669/pls_2021-2.

Bainbridge, Lisanne. 1987. »Ironies of Automation.« In: *New Technology and Human Error,* edited by Rasmussen, Jens, Keith Duncan, and Jacques Leplat. John Wiley, New York.

Benjamin, Ruha. 2019. »Race after Technology: Abolitionist Tools for the New Jim Code.« *Social Forces* 98 (4). DOI: https://doi.org/10.1093/sf/soz162.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock. 2018. »What Can Machines Learn, and What Does It Mean for Occupations and the Economy?« *AEA Papers and Proceedings* 108: 43–47. DOI: https://doi.org/10.1257/pandp.20181019.

Cherns, Albert. 1976. »The Principles of Sociotechnical Design.« *Human Relations* 29 (8): 783–92. DOI: https://doi.org/10.1177/001872677602900806.

Cherns, Albert. 1987. »Principles of Sociotechnical Design Revisted.« *Human Relations* 40 (3): 153–61. DOI: https://doi.org/10.1177/001872678704000303.

Deutsches Institut für Normung: Deutsche Normungsroadmap Künstliche Intelligenz. Ausgabe 2. 2023. (DIN-KI).

DIN EN 614–2:2008, Sicherheit von Maschinen – Ergonomische Gestaltungsgrundsätze – Teil 2: Wechselwirkungen zwischen der Gestaltung von Maschinen und den Arbeitsaufgaben; Deutsche Fassung EN 614–2:2000+A1:2008.

DIN EN 894–1:2009, Sicherheit von Maschinen – Ergonomische Anforderungen an die Gestaltung von Anzeigen und Stellteilen – Teil 1: Allgemeine Leitsätze für Benutzer-Interaktion mit Anzeigen und Stellteilen; Deutsche Fassung EN 894–1:1997+A1:2008.

DIN EN ISO 26800:2011, Ergonomie – Genereller Ansatz, Prinzipien und Konzepte (ISO 26800:2011); Deutsche Fassung EN ISO 26800:2011.

DIN EN ISO 9001:2015, Qualitätsmanagementsysteme – Anforderungen (ISO 9001:2015); Deutsche und Englische Fassung EN ISO 9001:2015.

DIN 69909 (alle Teile), Multiprojektmanagement – Management von Projektportfolios, Programmen und Projekten.

DIN EN ISO 6385:2016, Grundsätze der Ergonomie für die Gestaltung von Arbeitssystemen (ISO 6385:2016); Deutsche Fassung EN ISO 6385:2016.

DIN EN ISO 9241 (alle Teile), Ergonomie der Mensch-System-Interaktion.

DIN ISO 21500:2016, Leitlinien Projektmanagement (ISO 21500:2012).

DIN 69901 (alle Teile), Projektmanagement – Projektmanagementsysteme.

DIN EN ISO 9241–112:2017, Ergonomie der Mensch-System-Interaktion – Teil 112: Grundsätze der Informationsdarstellung (ISO 9241–112:2017); Deutsche Fassung EN ISO 9241–112:201.

DIN EN ISO 27500:2017, Die menschzentrierte Organisation – Zweck und allgemeine Grundsätze (ISO 27500:2016); Deutsche Fassung EN ISO 27500:2017.

DIN EN ISO 9241–11:2018, Ergonomie der Mensch-System-Interaktion – Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte (ISO 9241–11:2018); Deutsche Fassung EN ISO 9241–11:2018.

DIN SPEC 92001–1:2019, Künstliche Intelligenz – Life Cycle Prozesse und Qualitätsanforderungen – Teil 1: Qualitäts-Meta-Modell. Letzter Zugriff: 1. April 2023. https://www.beuth.de/de/technische-regel/din-spec-92001-1/303650673.

DIN SPEC 92001–2:2020, Artificial Intelligence – Life Cycle Processes and Quality Requirements – Part 2: Robustness.

DIN SPEC 92001–3, Künstliche Intelligenz – Life Cycle Prozesse und Qualitätsanforderungen – Teil 3: Explainability.

DIN EN ISO 9241–110:2020, Ergonomie der Mensch-System-Interaktion – Teil 110: Interaktionsprinzipien (ISO 9241–110:2020); Deutsche Fassung EN ISO 9241–110:2020.

DIN EN ISO 10075 (alle Teile), Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung.

Europäische Kommission. 2021. *Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union*. (EU-AI-Act). Accessed April 1, 2023. https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX :52021PC0206.

Fitts, Paul M. 1951. *Human engineering for an effective air-navigation and traffic-control system*. Washington, DC: National Research Council.

Hacker, Winfried, and Pierre Sachse. 2013. *Allgemeine Arbeitspsychologie: Psychische Regulation von Tätigkeiten*. Göttingen.

Jessica Heesen, Jörn Müller-Quade, Stefan Wrobel et al. 2021. *Kritikalität von KI-Systemen in ihren jeweiligen Anwendungskontexten – Ein notwendiger, aber nicht hinreichender Baustein für Vertrauenswürdigkeit*. Whitepaper aus der Plattform Lernende Systeme, München.

Herrmann, Thomas. 2020. *Socio-technical design of hybrid Intelligence systems- the case of predictive maintenance*. Accessed April 1, 2023. https://link.springer.com/chapter/10.1007/978-3-030-50334-5_20.

Höddinghaus, Miriam, Dominik Sondern, and Guido Hertel. 2021. »The Automation of Leadership Functions: Would People Trust Decision Algorithms?« *Computers in Human Behavior* 116 (March): 106635. DOI: https://doi.org/10.1016/j.chb.2020.106635.

Huchler, Norbert et al. 2020. *Kriterien für die Mensch-Maschine-Interaktion bei KI. Ansätze für die menschengerechte Gestaltung in der Arbeitswelt*. Accessed April 1, 2023. https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG2_Whitepaper2_220620.pdf.

IEEE 7000:2021, IEEE Standard Model Process for Addressing Ethical Concerns during System Design. Accessed April 1, 2023. https://standards.ieee.org/ieee/7000/6781/#.

IEEE 7001:2021, Standard for Transparency of Autonomous Systems.

IEEE 7002:2022, Standard for Data Privacy Process.

IEEE 7007:2021, Ontological Standard for Ethically driven Robotics and Automation Systems.

IEEE 7005:2021, Transparent Employer Data Governance.

ISO/IEC 29138–1:2018, Informationstechnik – Barrierefreie Benutzungsschnittstellen – Teil 1: Barrierefreiheitserfordernisse der Benutzer.

ISO/IEC 22989:2022, Informationstechnik – Künstliche Intelligenz – Konzepte und Terminologie der Künstlichen Intelligenz.

ISO/IEC DIS 23894:2022, Informationstechnik – Künstliche Intelligenz – Risikomanagement.

ISO/IEC 23053:2022, Framework für Systeme der Künstlichen Intelligenz (KI) basierend auf maschinellem Lernen (ML), 2022.

ISO/IEC DIS 25059:2022-07 – Entwurf, System- und Software-Engineering – Qualitätskriterien und Bewertung von Systemen und Softwareprodukten (SquaRE) – Qualitätsmodell für KI-System.

ISO/IEC DIS 42001, Information Technology – Artificial intelligence – Management system.

ISO/IEC TR 24028:2020, Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence.

ISO/IEC 5259 (alle Teile), Artificial intelligence – Data quality for analytics and machine learning (ML).

Kahneman, Daniel, Paul Slovic, and Amos Tversky. 1982 (1974). »Intuitive prediction: Biases and corrective procedures.« In: *Judgment under Uncertainty:*

Heuristics and Biases, edited by Kahneman, Daniel, Paul Slovic, and Amos Tversky. Cambridge: Cambridge University Press.

Moray, Neville. 2001. *Human and machines. Allocation of functions*. In: *People in Control: Human Factors in Control Room Design*, edited by Noyes, Janet M, and Matthew Bransby. London: Institution Of Electrical Engineers, 101–115.

Muhammad, Selma. 2022. *The Fairness Handbook*, Accessed April 1, 2023. https://www.amsterdamintelligence.com/resources/the-fairness-handbook.

Pentenrieder, Annelie and Jutta Weber. 2020. »Lucy Suchman.« In: *Technikanthropologie Handbuch Für Wissenschaft Und Studium*, edited by Heßler, Martina, Kevin Liggieri, and Nomos Verlagsgesellschaft. Baden-Baden Nomos, Edition Sigma, 215–224.

Puntschuh, Michael and Lajla Fetic. 2020. *Handreichung für die digitale Verwaltung. Algorithmische Assistenzsysteme gemeinwohlorientiert gestalten*. Bertelsmann Stiftung, Gütersloh. DOI: https://doi.org/10.11586/2020060.

Raisch, Sebastian, and Sebastian Krakowski. 2020. »Artificial Intelligence and Management: The Automation-Augmentation Paradox.« *Academy of Management Review*, February. DOI: https://doi.org/10.5465/2018.0072.

Rohde, Friederike et al. 2021. *Nachhaltigkeitskriterien für künstliche Intelligenz – Entwicklung eines Kriterien- und Indikatorensets für die Nachhaltigkeitsbewertung von KI-Systemen entlang des Lebenszyklus*. Accessed April 1, 2023. https://www.ioew.de/publikation/nachhaltigkeitskriterien_fuer_kuenstliche_intelligenz.

Schlick, Christopher, Ralph Bruder and Holger Luczak. 2010. *Arbeitswissenschaft*. 3. Auflage, Springer-Verlag Berlin Heidelberg.

Simonsen, Jesper, and Toni Robertson. 2013. *Routledge International Handbook of Participatory Design*. New York: Routledge.

Stowasser, Sascha and Oliver Suchy et al. (eds.). 2020. *Einführung von KI-Systemen in Unternehmen. Gestaltungsansätze für das Change-Management*. Whitepaper aus der Plattform Lernende Systeme, München.

Suchman, Lucille A. 2009. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge: Cambridge Univ. Pr.

Ulich, Eberhard. 2011. *Arbeitspsychologie*. Stuttgart.

Ulich, Eberhard. 2013. »Arbeitssysteme als Soziotechnische Systeme – eine Erinnerung.« *Journal Psychologie des Alltagshandelns/Psychology of Everyday Activity*, 6 (1), 4–12.

VDE, Bertelsmann Stiftung (Hg.). *From Principles to Practice – An interdisciplinary framework to operationalise AI ethics*. AI Ethics Impact Group, VDE Association for Electrical Electronic & Information Technologies e. V., Bertels-

mann Stiftung, 1–56, 2020. Accessed April 1, 2023. DOI: https://doi.org/10.11586/2020013.

VDI/VDE-MT 7100 – Entwurf, Lernförderliche Arbeitsgestaltung – Ziele, Nutzen, Begriffe.

VDI/VDE 3850–1:2014, Gebrauchstaugliche Gestaltung von Benutzungsschnittstellen für technische Anlagen – Konzepte, Prinzipien und grundsätzliche Empfehlungen.

VDE SPEC 90012:2022, VCIO based description of systems for AI trustworthiness characterization.

Zweig, Katharina A, Tobias Krafft, and Enno Park. 2021. *Sozioinformatik. Ein neuer Blick auf Informatik und Gesellschaft*. München.