

# Handreichung zur Gestaltung und Anwendung Künstlicher Intelligenz (KI)

## Gestaltungsgrundsätze für Entwickler\*innen, Beschäftigte und ihre Interessenvertretungen

### [1] **Umfassende Beteiligung und prozessbegleitende Mitbestimmung**

KI ist eine Schlüsseltechnologie des 21. Jahrhunderts mit weitreichenden Folgen für die Gesellschaft. Sie geht mit einer Transformation der Arbeit einher, bei der voraussichtlich viele Tätigkeiten automatisiert, andere tiefgreifend geändert, neue entstehen werden.

Um diesen Prozess sozial, menschengerecht und gemeinwohlorientiert, unter Wahrung der Interessen der Beschäftigten, mit ihrer Unterstützung, ihrem Wissen und ihren Erfahrungen zu gestalten, ist eine umfassende Beteiligung der Beschäftigten und ihrer Interessenvertretungen unabdingbar. Zu beachten ist dabei, dass bereits in frühen Stadien der Planung und Beschaffung weitreichende Weichenstellungen erfolgen, die Prozesse der Entwicklung und Einführung sich meist über längere Zeiträume erstrecken und mit der Produktivsetzung eines Systems, dessen Entwicklung nicht endet.

Die Auswirkungen auf die Arbeitsbedingungen sind vorab schwer abschätzbar, insbesondere nicht zu dem Zeitpunkt, zu dem Interessenvertretungen im Allgemeinen gebeten oder aufgefordert werden, ihre Zustimmung zur Einführung eines IT-Systems zu geben. In Anbetracht dessen sind frühzeitige Informationen im Planungsstadium sowie fortlaufende Informationen und eine weitgehende Mitbestimmung über den gesamten Projektverlauf und das Änderungsmanagement im weiteren Betrieb geboten.

Besonders zu beachten ist, dass die Anpassungsfähigkeit ein genuines Merkmal lernender KI-Systeme ist. Eine automatisierte Anpassung kann unwillkürlich im Zuge ihrer Anwendung oder gezielt erfolgen, um das System einer sich ständig wandelnden Umgebung anzupassen oder von Fehlern zu bereinigen und zu optimieren. Die maschinellen Lernverfahren bergen wiederum Gefahren, die Funktionstüchtigkeit des Systems zu untergraben, womöglich in einem Ausmaß, das einer verantwortungsbewussten Weiternutzung des Systems entgegensteht.

Der Einsatz agiler Methoden der Entwicklung und Anpassung von KI-Systemen wie maschineller Lernverfahren erfordert agile Verfahren der Beteiligung und Mitbestimmung. Es ist notwendig, dass sich die Mitbestimmung auf alle Phasen von KI-Vorhaben erstreckt und wesentliche Änderungen und Erweiterungen der Systeme einschließt.

## [2] Wahrung der Persönlichkeitsrechte

Bei der Gestaltung und Anwendung von KI-Systemen müssen die Persönlichkeitsrechte aller Betroffenen, insbesondere der an der Entwicklung und Anwendung beteiligten Beschäftigten, gewahrt werden. Dies schließt die Beachtung der Bestimmungen des Datenschutzes gemäß DSGVO und BDSG sowie der Bestimmungen des Arbeits- und Gesundheitsschutzes gemäß Arbeitsschutzgesetz und Arbeitsstättenverordnung ein. Durch technische und organisatorische Maßnahmen ist zu gewährleisten,

- ▷ dass die freie Entfaltung der Persönlichkeit gefördert und geschützt wird, insbesondere dass die menschliche Souveränität bei der Anwendung algorithmischer Entscheidungssysteme (ADMs<sup>1</sup>) gewahrt bleibt (Souveränitätsgebot) und die Aufgaben der Beschäftigten und Interoperabilität mit KI-Systemen menschengerecht gestaltet werden;
- ▷ dass unzulässige Leistungs- und Verhaltenskontrollen ausgeschlossen werden;
- ▷ dass natürliche Personen nicht einer voll- oder überwiegend<sup>2</sup> automatisierten Entscheidung unterworfen werden, die rechtliche Wirkung für sie entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt;
- ▷ dass die Verarbeitung besonderer Kategorien personenbezogener Daten gemäß Artikel 9 DSGVO nur zulässig ist, wenn die besonderen Schutzbedürfnisse für diese Datenkategorien unter Berücksichtigung der besonderen Merkmale der eingesetzten KI-Systeme, insbesondere Risiken von Fehlschlüssen und Diskriminierungen, beachtet werden; die Verarbeitung besonderer Kategorien personenbezogener Daten erfordert grundsätzlich eine Technologiefolgenabschätzung (TFA, siehe [Technologiefolgenabschätzung](#)) und die Umsetzung daraus hervorgehender Auflagen;
- ▷ dass die Arbeit gesundheitsgerecht gestaltet wird, insbesondere die erforderlichen Sicherheitsvorkehrungen im Umgang mit Robotern getroffen werden und gesundheitsgefährdende psychische Belastungen aufgrund nicht oder schwerlich nachvollziehbarer und/oder risiko- und folgenreicher Entscheidungen von ADMs ausgeschlossen werden.<sup>3</sup>

---

<sup>1</sup>ADM: Algorithmic Decision Making System.

<sup>2</sup>Die Erweiterung auf „überwiegend automatisierte Entscheidungen“ geht über die Vorschrift von Artikel 22 Abs. 1 DSGVO hinaus. Sie betrifft Entscheidungen, die in einem so erheblichen Maß automatisiert erfolgen oder vorbereitet werden, dass die Entscheidung durch eine natürliche Person überwiegend formaler Natur ist und eine angemessene Sacherwägung und Prüfung der vom System vorgeschlagenen Entscheidung aufgrund der Umstände, insbesondere verfügbarer Zeit, Information und Kompetenz, nachweislich oder mutmaßlich nicht erfolgt ist und praktisch schwerlich erfolgen kann.

<sup>3</sup>Die gesundheitsgerechte Gestaltung der Arbeit ist ein Gebot des Arbeits- und Gesundheitsschutzes. Maßgebliche Rechtsgrundlagen hierfür sind das Arbeitsschutzgesetz (ArbSchG) und die Arbeitsstättenverordnung (ArbStättV). Der Arbeitgeber ist zur Durchführung einer Gefährdungsbeurteilung, zur Umsetzung daraus hervorgehender erforderlicher Maßnahmen sowie zur Dokumentation und Wirksamkeitskontrolle dessen verpflichtet (siehe §§ 3, 4, 5 ArbSchG, §§ 3, 3a ArbStättV). Die Arbeit mit KI-Systemen bringt spezifische körperliche und psychische Gefährdungen mit sich, die im Rahmen der Gefährdungsbeurteilung zu ermitteln sind.

Die sachgerechte Planung und Umsetzung der Maßnahmen ist durch geeignete Maßnahmen der Beteiligung und Mitbestimmung (siehe [Umfassende Beteiligung und prozessbegleitende Mitbestimmung](#)) sowie der Technologiefolgenabschätzung (siehe [Technologiefolgenabschätzung](#)) und Evaluation (siehe [Evaluation](#)) sicherzustellen.

### [3] [Transparenz und Nachvollziehbarkeit](#)

KI-Systeme müssen einschlägigen Richtlinien der Transparenz und Nachvollziehbarkeit genügen (Transparenzgebot<sup>4</sup>).

- ▷ KI-Systeme mit personenbezogenen Daten müssen diese „auf rechtmäßige, nach Treu und Glauben und in einer für die betroffene Person nachvollziehbaren Weise“ (Art. 5 Abs. 1 lit. a DSGVO) verarbeiten.
- ▷ Für KI-Systeme muss eine Kennzeichnungspflicht eingeführt werden, die die Anwender- respektive Verbraucher\*innen über den Einsatz von KI-Methoden angemessen informiert.<sup>5</sup>
- ▷ KI-Systeme müssen grundsätzlich sachgerecht dokumentiert und so gestaltet werden, dass die von ihnen generierten Schlussfolgerungen/Entscheidungen nachvollziehbar sind (Transparenz-by-Design).<sup>6</sup> Dieser Grundsatz gilt uneingeschränkt für grundrechtssensitive Anwendungen oder Anwendungen, die erhebliche gesundheitliche, soziale oder monetäre Risiken bergen.
- ▷ Die Anforderungen an die Dokumentation sowie Transparenz und Nachvollziehbarkeit sind bezüglich der Anwendungsgebiete und Funktionen der Systeme sowie den damit verbundenen Risiken zu differenzieren.

Technische Verfahren zur Umsetzung der Anforderungen hängen von der Architektur der Systeme respektive den zugrundeliegenden Algorithmen (wie insb. Neuronale Netze/deep learning, Regel- und Logiksysteme, Bayessche Netze, Fuzzy-Systeme) ab.

---

<sup>4</sup>Das hier geforderte Transparenzgebot inkludiert das in der DSGVO verankerte Recht, geht aber zugleich, wie die folgenden Ausführungen verdeutlichen, über es hinaus. Es ist nicht primär und schon gar nicht einzig im informationellen Selbstbestimmungsrecht begründet und beschränkt sich nicht darauf, Risiken des Datenschutzes zu begegnen.

<sup>5</sup>Ein typischer Anwendungsfall hierfür wären Chatbots.

<sup>6</sup>Erklärungskomponenten waren von Beginn an ein integraler Bestandteil wissensbasierter Systeme oder Expertensysteme. Für regelbasierte Systeme wurden sehr elaborierte Erklärungskomponenten bis hin zu „intelligenten Tutorsystemen“ entwickelt. Dies belegen bspw. die Systeme Mycin, Neomycin und Guidon. Die in der symbolorientierten KI entwickelten Ansätze zur Erklärungsfunktionalität sind nicht einfach übertragbar auf andere Verfahren wie künstliche neuronale Netze, insb. deep learning. Sie sind aber auch nicht obsolet, weil symbolorientierte Verfahren und Wissensrepräsentationsformalismen weiter zum Einsatz kommen (siehe bspw. Watson) und Maßstäbe zur Erklärungsfunktionalität gesetzt haben. Für künstliche neuronale Netze wurden und werden andere Verfahren entwickelt und erforscht. Dazu gehören insbesondere auch eigenständige Erklärungssysteme für ADMs wie bspw. LIME (Local Interpretable Model-Agnostic Explanations). Die Forschungen auf diesem Gebiet sollten intensiviert werden und die Gebrauchstauglichkeit derartiger Systeme sollte genauer untersucht werden, um ihre Erklärungs- und Kontrollfunktionalität in der Praxis besser beurteilen und verbindliche Standards für derartige Systeme entwickeln zu können.

Umfang und Darstellungsweise von Erklärungen sind zudem nach Stakeholdern (insb. betroffene Person gemäß Art. 4 Abs. 1 DSGVO, Entwickler, Anwender, Verbraucher, Interessenvertretungen, Öffentlichkeit) zu differenzieren.

Die Dokumentation muss mindestens Auskunft über die Systemarchitektur, insbesondere der verwandten Algorithmen, sowie die Wissens- oder Datenquellen bieten. Die Ausführlichkeit und Genauigkeit dieser und weiterer Informationen hängt vom Anwendungsfall ab.

Die Dokumentation muss grundsätzlich Änderungen der Wissens- oder Datenbasis, insbesondere Anreicherungen und Transformationen von Trainingsdaten von KI-Systemen mit maschinellen Lernverfahren, einschließen und revisions sicher sein.

- ▷ KI-Systeme, insbesondere ADMs, die auf Verfahren maschinellen Lernens basieren, müssen nachweislich einschlägigen Anforderungen an die Qualität der zugrundeliegenden Daten genügen.

Informationen über die Daten, die Datenquellen und datenerhebenden Stellen sowie das Datenqualitätsmanagement sind ein wichtiger Bestandteil der Kontrolle von KI-Systemen. Sie sind als Teil der Information über die involvierte Logik algorithmischer Entscheidungssysteme zu behandeln.

Im Rahmen des Datenqualitätsmanagements sind alle erforderlichen Maßnahmen zur Gewährleistung einer hohen Datenqualität einschließlich der Anwendung geeigneter statistischer Prüfverfahren zu ergreifen und Informationen bereitzustellen, die etwaige Abhängigkeiten von besonderen Nutzungskontexten und Verzerrungen bei der Datenerhebung erkennen lassen.

- ▷ Es sollten technische Regeln für KI-Systeme erstellt werden, die unter Beachtung der verschiedenen Transparenz-Anforderungen und Systemarchitekturen verbindliche Richtlinien vorgeben, die zur Dokumentation sowie zur Gestaltung und Prüfung dieser Systeme einschließlich diesbezüglicher Prüfwerkzeuge verwandt werden können.
- ▷ Anforderungen an die Transparenz von KI-Systemen schließen nicht zuletzt klare Festlegungen von Verantwortlichkeiten sowie die Klärung von Haftungsfragen ein.

Insbesondere ist bei der Anwendung von KI-Systemen über die Bestimmungen zu dem Verantwortlichen oder den gemeinsamen Verantwortlichen bezüglich der Datenverarbeitung gemäß DSGVO (Artikel 24ff.) eine Klärung der Abgrenzung von Hersteller- und Betreiberhaftung erforderlich.

#### [4] Gebrauchstauglichkeit von KI-Systemen

Gebrauchstauglichkeit gemäß ISO 9241-11 ist ein Gebot der effizienten und gesundheitsgerechten Gestaltung und Anwendung von IT-Systemen. Den Grundsätzen und Richtlinien der Gebrauchstaug-

lichkeit kommt eine maßgebliche Funktion bei der menschengerechten Gestaltung von KI-Systemen zu. Gebrauchstauglichkeit ist unabdingbar für gute digitale Arbeit mit KI. Ihr muss in Planung, Entwicklung, Beschaffung, Erprobung, Anpassung und Betrieb von KI-Systemen hohe Priorität eingeräumt werden.

Die Maße und Richtlinien der Gebrauchstauglichkeit sind anzuwenden und zu konkretisieren für KI-Systeme und ihre jeweiligen Nutzungskontexte. Die Dialoggestaltung von KI-Systemen hat vor allem so zu erfolgen, dass die Kontrolle der Systeme, ihre Überwachung und der Nachvollzug von ihnen generierter Entscheidungen und sonstiger Schlussfolgerungen und Resultate möglichst einfach und effizient erfolgen kann.

Dies gilt insbesondere für Richtlinien der Benutzerführung und Informationsdarstellung sowie für die Grundsätze der Dialoggestaltung. So implizieren die Grundsätze der Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Fehlertoleranz und Lernförderlichkeit gemäß ISO 9241-110 Anforderungen an eine transparente und nachvollziehbare Gestaltung von KI-Systemen wie umgekehrt das Transparenzgebot (siehe **Transparenz und Nachvollziehbarkeit**) eine Konkretisierung dieser Grundsätze im Hinblick auf die Merkmale von KI-Systemen und Bestimmungen des Transparenzgebots ermöglicht und erfordert. Analoges gilt auch für den Grundsatz der Steuerbarkeit in Relation zum Souveränitätsgebot (siehe **Wahrung der Persönlichkeitsrechte**).

Anforderungen an die Gebrauchstauglichkeit betreffen sowohl die KI-Systeme selbst wie separate Werkzeuge, die der Überwachung, Überprüfung und Dokumentation von KI-Systemen, insbesondere der Bereitstellung von Erklärungen dienen.<sup>7</sup>

## [5] Technologiefolgenabschätzung

- ▷ Die KI-TFA schließt die Prüfung der Erforderlichkeit einer Datenschutz-Folgenabschätzung (DSFA) gemäß Artikel 35 DSGVO und ggf. ihre Durchführung ein.
- ▷ In der KI-TFA sind alle wesentlichen gesellschaftlichen Auswirkungen unter besonderer Beachtung der Grundsätze zur Wahrung der Persönlichkeitsrechte (siehe **Wahrung der Persönlichkeitsrechte**) zu berücksichtigen. Insbesondere sind Folgen für die Wahrung menschlicher Entscheidungsfreiheit und sozialer Diskriminierungsfreiheit zu untersuchen.
- ▷ Die KI-TFA schließt die Prüfung der Erforderlichkeit einer Datenschutz-Folgenabschätzung (DSFA) gemäß Artikel 35 DSGVO und ggf. ihre Durchführung ein.

---

<sup>7</sup>Dies betrifft vor allem ADMs auf Basis maschineller Lernverfahren (insbesondere deep learning). Dabei geht es um Verfahren und Applikationen, die dazu dienen, Entscheidungen dieser Systeme nachzuvollziehen und statistisch zu prüfen und so beispielsweise eine „antrainierte“ Unfairness oder Diskriminierung aufzudecken. Neben solchen Prüfsystemen (wie LIME) sind allgemeine Toolkits wie Algoneer (<https://algoneer.org>) zu betrachten, die sich solcher Analysesysteme bedienen.

Dabei sind einschlägige Richtlinien wie ISO/IEC 29134<sup>8</sup> zu berücksichtigen.

An der DSFA für KI-Systeme (KI-DSFA) ist ein Assessor zu beteiligen, der über einschlägige rechtliche und technische Kenntnisse verfügt; die technischen Kenntnisse müssen mindestens Grundkenntnisse der eingesetzten KI-Verfahren einschließen.

Im Rahmen der KI-DSFA hat grundsätzlich eine Konsultation der interessierten Parteien, wie insbesondere Interessenvertretungen, zu erfolgen.<sup>9</sup>

- ▷ Die KI-TFA für Systeme, die im Rahmen eines Beschäftigtenverhältnisses zum Einsatz kommen, muss eine Folgeabschätzung bezüglich der Auswirkungen auf die Arbeitsbedingungen beinhalten. Dazu gehören insbesondere Auswirkungen auf die Personalentwicklung und Bestimmungen für die Datenverarbeitung im Beschäftigtenverhältnis (gemäß Artikel 35 und 88 DSGVO sowie § 26 BDSG). Insofern die Systeme als Arbeitsmittel eingesetzt werden, ist grundsätzlich eine prospektive Gefährdungsbeurteilung unter Beachtung einschlägiger Richtlinien des Arbeits- und Gesundheitsschutzes durchzuführen.

Hierzu gehören: „Technischen Regeln für Arbeitsstätten“ (ASR)<sup>10</sup> und die internationale Norm zur „Mensch-Computer-Interaktion“ (ISO 9241). Bezüglich der psychischen Gefährdungsbeurteilung sind insb. die „Empfehlungen zur Umsetzung der Gefährdungsbeurteilung psychischer Belastung“ der GDA (Gemeinsame Deutsche Arbeitsschutzstrategie) zu beachten.

## [6] Evaluation

KI-Systeme sind grundsätzlich im Hinblick auf ihre gesellschaftlichen Auswirkungen, insbesondere der Wahrung der Persönlichkeitsrechte (siehe [Wahrung der Persönlichkeitsrechte](#)), zu evaluieren.

Ausnahmen betreffen Systeme, die nach den Befunden der KI-TFA voraussichtlich keine erheblichen Auswirkungen haben und aus deren praktischer Anwendung kein begründeter Verdacht auf erhebliche Auswirkungen erwachsen ist.

Methoden, Umfang, Zeitpunkte und Zielsetzungen der Evaluation sind sachgerecht unter besonderer Berücksichtigung der Anwendungsgebiete, der verwandten KI-Methoden, der Risikoeinstufung und Folgenabschätzung sowie diesbezüglicher Standards festzulegen.<sup>11</sup>

<sup>8</sup>Der Titel der Norm lautet: „Information technology – Security techniques – Guidelines for privacy impact assessment“.

<sup>9</sup>Diese Anforderung erweitert die Bestimmung von Artikel 35 Abs. 9 DSGVO in spezifischer Hinsicht und verbindlicher Form.

<sup>10</sup>Die Arbeitsstättenrichtlinien (ASR) werden vom Ausschuss für Arbeitsstätten (ASTA) erarbeitet und treten nach der Veröffentlichung durch das Bundesministerium für Arbeit und Soziales in Kraft.

<sup>11</sup>Zu denken ist hierbei insbesondere an technische Regelwerke, Checklisten, allgemeine Leitlinien, Verfahren bzw. Vorgehensweisen, beispielsweise an Gutachten, moderierte Workshops, Audits.

Die Evaluation sollte der Überprüfung vorab festgelegter Erfolgskriterien dienen und dabei so offen gestaltet sein, dass neue Erfahrungen und unerwartete Auswirkungen erfasst und bewertet werden können.

Konzeption und Zeitpunkte der Durchführung sollen darauf ausgerichtet sein, rechtzeitig und bedarfsgerecht organisatorische oder technische Maßnahmen treffen zu können, die zur Optimierung des Verfahrens dienen oder notfalls Konsequenzen im Hinblick auf einen ineffizienten oder zu risikohaften Weiterbetrieb des Systems zu ziehen erlauben.

Eine Evaluation sollte grundsätzlich im Rahmen des Pilotbetriebs erfolgen, um erhebliche Mängel im Nutzungskontext frühzeitig im Licht der Praxisanforderungen unter Mitwirkung der Beschäftigten erkennen zu können und die Interessenvertretungen darin zu unterstützen, ihre Beteiligungsrechte zur Gewährleistung einer menschengerechten Gestaltung der Systeme effektiv wahrnehmen zu können.

Sie sollte überdies bedarfsweise im Laufe des weiteren Betriebs durchgeführt werden. Eine Evaluation kann auch vorab auf Grundanforderungen, Bausteine oder Teile des Systems angezeigt sein. Dies gilt vor allem bei der Verwendung vortrainierter Netze, insbesondere auch im Hinblick auf potentiell sozial diskriminierende Verzerrungen in Trainingsdatensätzen für Deep-Learning-Systeme oder andere künstliche neuronale Netze.<sup>12</sup>

Bei der Evaluation sind Änderungen der Systeme und ihres Nutzungskontexts zu beachten. Insbesondere bei Systemen, die maschinelle Lernverfahren einsetzen und aufgrund dessen einer ständigen Adaption unterliegen, sollten grundsätzlich anwendungsbegleitende Evaluationen unter Verwendung automatisierter Prüftools durchgeführt werden.

## **[7] Qualifizierung – Handlungsfähigkeit im Umgang mit KI-Systemen**

Der sachgerechte Umgang mit KI-Systemen stellt besondere Anforderungen an die Qualifikation aller bei der Planung, Entwicklung, Beschaffung, Prüfung, Einführung und Anwendung Beteiligten.

Die Qualifikationsanforderungen können von Fall zu Fall höchst unterschiedlich ausfallen, insbesondere in Abhängigkeit von den verwandten KI-Methoden, den Anwendungsgebieten, den Risiken sowie den Aufgaben, Funktionen oder Rollen der Beteiligten. Um verantwortungsvoll die Möglichkeiten der KI nutzen und ihren Gefahren begegnen zu können ist eine hohe Qualifikation auf vielen

---

<sup>12</sup>Besonders zu beachten sind in diesem Zusammenhang Vorhersagesysteme (predictive analytics), die beispielsweise im Bereich der Verbrechensbekämpfung oder des Personalmanagements, insbesondere der Personalrekrutierung und Personalentwicklung, eingesetzt werden. Im Hinblick auf potentielle Verzerrungen bei sprach-/textverarbeitenden KI-Systemen sei hier auch auf die Verwendung von Wortvektoren (word vectors, word embedding) verwiesen. Ob und inwieweit die dazu verwandten Datenquellen „fair“ und diskriminierungsfrei sind, erscheint fraglich. Etwaige Verzerrungen in den Datenquellen können sich in die aus ihnen berechneten Wortvektoren und die sich daraus speisenden künstlichen neuronalen Netze fortpflanzen.

Gebieten der Gestaltung, Steuerung und Anwendung erforderlich. Bei Planung und Umsetzung von KI-Vorhaben muss gewährleistet werden, dass alle Beteiligten über die erforderlichen Qualifikationen verfügen, um im Rahmen ihrer Aufgabe zu einer menschengerechten Gestaltung und Anwendung von KI-Systemen beitragen zu können. Insbesondere sind folgende Qualifikationsanforderungen zu erfüllen:

- ▷ alle Beteiligten eines KI-Vorhabens sollten mindestens über ein Grundverständnis der Funktionsweise der infrage stehenden Systeme sowie mit ihnen verbundene Potentiale und Risiken verfügen, das sie befähigt ihre Aufgaben im Rahmen des Vorhabens kompetent zu erfüllen;
- ▷ Entwicklungsteams von KI-Systemen sollten neben einschlägigen Kenntnissen zu den angewandten Programmiermethoden, Entwicklungs- und Prüfwerkzeugen, einschlägige Kenntnisse über Grundsätze einer menschengerechten Gestaltung, einschließlich diesbezüglicher ethischer Leitlinien, einschlägiger Bestimmungen des Datenschutzes und ergonomischer Richtlinien, verfügen;
- ▷ Beschäftigte, die ein KI-System, insbesondere ADM, als Fachanwendung im Rahmen ihrer Aufgabenerfüllung einsetzen, sind umfassend zu schulen. Sie sollten über allgemeine Kenntnisse und Kompetenzen in der Anwendung des Systems hinaus, vertiefte Kenntnisse im kritischen und souveränen Umgang mit ihm haben.<sup>13</sup>

Der Einsatz Künstlicher Intelligenz wird Arbeitswelt und Gesellschaft grundlegend verändern. Die hier aufgeführten Leitlinien und Gestaltungsgrundsätze sollen dazu beitragen, dass dieser Transformationsprozess demokratisch und gemeinwohlorientiert vollzogen und mit ihm ein wichtiges Prinzip etabliert wird: Gute Arbeit by design.

---

**Impressum:** ver.di | Vereinte Dienstleistungsgewerkschaft

Michael Fischer, Politik und Planung  
Paula-Thiede-Ufer 10, 10179 Berlin  
verdi.de | [info@verdi.de](mailto:info@verdi.de)  
November 2020  
[www.verdi.de/themen/digitalisierung](http://www.verdi.de/themen/digitalisierung)

Diese Leitlinien zur Gestaltung und Anwendung Künstlicher Intelligenz wurden erstellt in Kooperation mit der tbs berlin GmbH

Dr. Benno Biewer  
tbs berlin GmbH  
Kapweg 4  
13405 Berlin  
[www.tbs-berlin.de](http://www.tbs-berlin.de) | [b.biewer@tbs-berlin.de](mailto:b.biewer@tbs-berlin.de)

---

<sup>13</sup>Dies schließt ein Grundverständnis technologisch bedingter grundsätzlicher oder potentieller Grenzen und Schwachstellen des Systems (bspw. Fehler- oder Diskriminierungswahrscheinlichkeit in Abhängigkeit von Trainingsdaten bei maschinellem Lernen), Fähigkeiten zum Infragestellen, zum Nachvollzug und zur Prüfung von Entscheidungen des Systems mit Hilfe systemeigener oder externer Prüfwerkzeuge ein; es schließt ferner den kompetenten Umgang mit fragwürdigen oder fehlerhaften Schlussfolgerungen und die Fähigkeit mit ein, eigene Erfahrungen und Anforderungen an die Wartung und Weiterentwicklung des Systems in das Änderungsmanagement einbringen zu können.