

Vogel-Adham, Elke; Ritzmann, Susanne; Blanc, Berit; Hochbauer, Monica; Reichow, Insa
**Sozio-ethische Aspekte KI-gestützter Bildungstechnologien. Empfehlungen
eines Expert_innen-Workshops. Ein Dossier im Rahmen des
Innovationswettbewerbs INVITE**

Berlin 2023, 22 S.



Quellenangabe/ Reference:

Vogel-Adham, Elke; Ritzmann, Susanne; Blanc, Berit; Hochbauer, Monica; Reichow, Insa:
Sozio-ethische Aspekte KI-gestützter Bildungstechnologien. Empfehlungen eines
Expert_innen-Workshops. Ein Dossier im Rahmen des Innovationswettbewerbs INVITE. Berlin 2023, 22
S. - URN: urn:nbn:de:0111-pedocs-261102 - DOI: 10.25656/01:26110

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-261102>

<https://doi.org/10.25656/01:26110>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz:
<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.de> - Sie dürfen das
Werk bzw. den Inhalt unter folgenden Bedingungen vervielfältigen, verbreiten
und öffentlich zugänglich machen: Sie müssen den Namen des
Autors/Rechteinhabers in der von ihm festgelegten Weise nennen. Dieses
Werk bzw. dieser Inhalt darf nicht für kommerzielle Zwecke verwendet
werden und es darf nicht bearbeitet, abgewandelt oder in anderer Weise
verändert werden.

Mit der Verwendung dieses Dokuments erkennen Sie die
Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License:
<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en> - You may copy,
distribute and transmit, adapt or exhibit the work in the public as long as you
attribute the work in the manner specified by the author or licensor. You are
not allowed to make commercial use of the work or its contents. You are not
allowed to alter, transform, or change this work in any other way.

By using this particular document, you accept the above-stated conditions of
use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Elke Vogel-Adham, Susanne Ritzmann, Berit Blanc, Monica Hochbauer & Insa Reichow

Sozio-ethische Aspekte KI-gestützter Bildungstechnologien Empfehlungen eines Expert_innen-Workshops

Ein Dossier im Rahmen des Innovationswettbewerbs INVITE

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



Bundesinstitut für
Berufsbildung

Impressum

Elke Vogel-Adham, Dr. Susanne Ritzmann

VDI/VDE Innovation + Technik GmbH
Steinplatz 1
10623 Berlin

Gemeinsame Kontaktadresse der Autorinnen: digitalbegleitung@vdivde-it.de

Dr. Berit Blanc, Dr. Insa Reichow

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)
Alt-Moabit 91c
10559 Berlin

Monica Hochbauer

mmb Institut GmbH
Folkwangstraße 1
45128 Essen

Gemeinsame Kontaktadresse der Autorinnen von DFKI und mmb: invite@mmb-institut.de

Dieses Dossier ist im Rahmen des Innovationswettbewerbs INVITE im Projekt „INVITE-Meta“ in Zusammenarbeit mit der INVITE-Digitalbegleitung von VDI/VDE IT entstanden. Elke Vogel-Adham und Dr. Susanne Ritzmann arbeiten bzw. arbeiteten zum Zeitpunkt der Dossiererstellung im Team der Digitalbegleitung. Dr. Berit Blanc, Monica Hochbauer und Dr. Insa Reichow sind Teil des Projektteams von INVITE-Meta.

Danksagung

Wir möchten uns herzlich bei allen Personen bedanken, die sich an der Ausarbeitung und Kommentierung dieses Dossiers beteiligt haben. Dazu gehören zum einen die Teilnehmer*innen des Expert_innen-Workshops, der die Grundlage des Dossiers bildet: Lajla Fetic (Bertelsmann Stiftung), Christoph Hohenberger (KI-Bundesverband), Prof. Claudia Lemke (HWR Berlin), Paola Lopez (Universität Wien), Prof. Tobias Röhl (Pädagogische Hochschule Zürich), Sebastian Straub (VDI/VDE-IT) und Dorothea Winter (DFKI).

Zum anderen danken wir den weiteren Mitarbeitern und Mitarbeiterinnen des Projekts INVITE-Meta: Dr. Lutz Goertz, Katja Buntins, Dr. Ulrich Schmid, Dr. Faisal Rashid, Dr. Christian Kellermann und Prof. Dr. Niels Pinkwart sowie der INVITE-Digitalbegleitung.

Zitiervorschlag:

Vogel-Adham, E., Ritzmann, S., Blanc, B., Hochbauer, M., Reichow, I. (2023). Sozio-ethische Aspekte KI-gestützter Bildungstechnologien. Ein Dossier im Rahmen des Innovationswettbewerbs INVITE. Berlin.

1. Auflage 2023
Februar 2023



Herausgeber

Deutsches Forschungszentrum
für Künstliche Intelligenz GmbH
Alt-Moabit 91c
10559 Berlin

CC Lizenz

Dieses Werk ist lizenziert unter einer CC BY-NC-ND 4.0 Lizenz (Namensnennung – Keine kommerzielle Nutzung – Keine Bearbeitung – 4.0 International). Weitere Informationen finden Sie auf der Creative-Commons-Webseite: <https://creativecommons.org/licenses/>

Inhaltsverzeichnis

Kurzzusammenfassung	4
1 Einführung.....	5
1.1 Rechtliche, technische und ethische Unbedenklichkeit KI-gestützter Bildungstechnologien.....	5
1.2 Expert_innen-Workshop.....	6
2 KI-gestützte Bildungstechnologien – Drei Anwendungsfälle basierend auf INVITE- Entwicklungen und Expert_innenfeedback.....	6
2.1 Anwendungsfall 1: Auswahl passender Weiterbildungsangebote.....	6
2.2 Anwendungsfall 2: Begleitung des Weiterbildungsprozesses.....	7
2.3 Anwendungsfall 3: Einsicht in die Lernaktivitäten der Beschäftigten.....	8
3 Übergreifende Empfehlungen	9
4 Weiterführende Ressourcen zur Umsetzung ethischer und vertrauenswürdiger KI- gestützter Bildungstechnologien.....	12
4.1 Ethisches Handeln im Projekt	12
4.2 Rechtliche Grundlagen	13
4.3 KI-Normen und -Standards	16
4.4 Zertifizierungen	17
4.5 Leitfäden, Handreichungen und Tools.....	17
4.6 Selbstverpflichtung.....	19
5 Literaturverzeichnis	21

Kurzzusammenfassung

An wen richtet sich dieses Dossier?

Dieses Dossier richtet sich an Entwickler_innen und andere Personen, die mit der konkreten Entwicklung und Implementierung von KI-gestützten Systemen im Bildungsbereich befasst sind.

Auf welchen Quellen basiert dieses Dossier?

Die in diesem Dossier beschriebenen Erkenntnisse und Empfehlungen basieren auf den Diskussionen eines Expert_innenworkshops zur europäischen KI-Verordnung (Europäische Kommission 2021) und zum Thema Ethik in der Entwicklung KI-gestützter Bildungstechnologien. Zu explizit von den Expert_innen genannter Literatur wurde im Nachgang recherchiert.

Was steht in diesem Dossier?

Das vorliegende Dossier unterscheidet zunächst die rechtliche, technische sowie ethische Unbedenklichkeit von KI-gestützten Bildungstechnologien (siehe Kapitel 1). Die nachfolgenden Kapitel konzentrieren sich auf die ethische Unbedenklichkeit.

Anhand dreier Anwendungsfälle wird im zweiten Kapitel konkret dargestellt, worauf bei der Entwicklung und Implementierung dieser Beispiele geachtet werden sollte. Anschließend beschreibt Kapitel 3 übergreifende Handlungsempfehlungen, die über die vorangegangenen Fallbeispiele hinausgehen. Das abschließende Kapitel 4 geht auf weiterführende Ressourcen zur Thematik ein.

Insgesamt soll das vorliegende Dossier damit einen praktischen Einstieg bieten, um sozio-ethische Aspekte bei der Entwicklung und Implementierung von KI-gestützten Systemen in der beruflichen Bildung zu berücksichtigen.

Was steht nicht in diesem Dossier?

Das vorliegende Dossier enthält keine Empfehlungen zur Anwendung der KI-Verordnung (rechtliche Unbedenklichkeit). Hierzu entsteht parallel zu dieser Publikation eine Veröffentlichung der Digitalbegleitung im Rahmen des INVITE-Wettbewerbs.

1 Einführung

1.1 Rechtliche, technische und ethische Unbedenklichkeit KI-gestützter Bildungstechnologien

KI-gestützte Technologien können für unterschiedliche Bildungszwecke und -kontexte eingesetzt werden. In digitalen Lernumgebungen unterstützen sie beispielsweise Lernprozesse, indem sie Angebote zusammenstellen, die genau zum Kenntnisstand und zu den individuellen Lernzielen passen. Im Bereich der beruflichen Aus- und Weiterbildung ist eine Vielzahl weiterer Einsatzzwecke KI-gestützter Bildungstechnologien denkbar. Das Bundesministerium für Bildung und Forschung (BMBF) fördert über den Innovationswettbewerb INVITE¹ zwischen 2021 und 2025 mit insgesamt 88 Mio. Euro 35 Projekte, in denen in unterschiedlichem Ausmaß Potenziale KI-gestützter Technologien für den Bereich der beruflichen Aus- und Weiterbildung ausgeschöpft werden.

Gerade im sensiblen Bereich der Bildung ist es dabei elementar, auf die Entwicklung und Implementierung vertrauenswürdiger, unbedenklicher KI-Systeme hinzuwirken. Bei der Prüfung auf Unbedenklichkeit von KI-Systemen ist es hier empfehlenswert, eine Unterscheidung in rechtliche, ethische/soziale und technische Unbedenklichkeit vorzunehmen und diese von Anfang an bei der Entwicklung von Systemen zu berücksichtigen. Die **rechtliche Unbedenklichkeit** bzw. die Rechtskonformität eines KI-Systems, liegt vor, wenn die Entwicklung und die Inbetriebnahme der KI im Einklang mit den bestehenden gesetzlichen Vorgaben stehen. Hier sind je nach Branche oder Einsatzzweck verschiedene Regelungsebenen zu betrachten. Mit der KI-Verordnung entscheidet sich der Gesetzgeber für einen horizontalen Regulierungsansatz. Das bedeutet, dass die Vorgaben der Verordnung unabhängig von einem bestimmten Sektor oder einer bestimmten Branche, gelten. Unter welchen Bedingungen und wie genau die KI-Verordnung und ihre Anforderungen für KI-gestützte Bildungstechnologien eine Rolle spielen wird, kann der aktualisierten Handreichung zum Verordnungstext entnommen werden².

Während eine Missachtung der Rechtslage natürlich unangenehme und handfeste Konsequenzen nach sich ziehen kann (z. B. Bußgelder, Marktverbot), so ist das Gewährleisten **technischer Unbedenklichkeit** von KI-Systemen vornehmlich eine Frage der Umsetzung und Implementierung. Hierbei gilt es, Systeme zum einen robust und sicher für den Betrieb³ zu entwickeln sowie zum anderen die Qualität und Fehlerfreiheit der verwendeten KI-Modelle und Algorithmen sicherzustellen. Den wohl schwierigsten Aspekt der Prüfung stellt die ethische Unbedenklichkeit dar. Eine Prüfung auf **ethische Unbedenklichkeit** umfasst dabei gesellschaftliche, soziale, kulturelle und wertebasierte Aspekte, die zwar (noch) nicht Teil der offi-

¹ Mehr Informationen zum Innovationswettbewerb INVITE unter: <https://www.bibb.de/de/120851.php>

² Die KI-Verordnung wird voraussichtlich im ersten Quartal 2023 verabschiedet. Nach ihrer Annahme wird die Verordnung 20 Tage nach ihrer Veröffentlichung im Amtsblatt der Europäischen Union in Kraft treten. Nach einer Übergangszeit von 36 Monaten wird die Verordnung (voraussichtlich) 2026 unmittelbar und ohne weiteren nationalen Umsetzungsakt EU-weit anwendbar sein. In dem INVITE-internen Technologiepapier der Digitalbegleitung zur Regulierung von Künstlicher Intelligenz und die Auswirkungen der geplanten KI-Verordnung auf Bildungsplattformen gibt Straub 2022 einen Überblick der Bedingungen und Anforderungen der geplanten KI-Verordnung. Eine Aktualisierung des Technologiepapiers im Jahr 2023 nach Inkrafttreten der KI-Verordnung ist geplant.

³ Orientierung bieten hier: Technische Richtlinie TR-03147 (Bundesamt für Sicherheit in der Informationstechnik 2021), BSI (Technische Richtlinie TR-03107-1, Elektronische Identitäten und Vertrauensdienste im E-Government, 2019) bzw. ISO 27001/ISO27002, ISO 27701 oder Standards wie der BSI-Grundschutz.

ziellen Rechtsprechung, jedoch zentral für eine zuverlässige und tragfähige Ausgestaltung technischer Systeme sind. Die Spannbreite ethischer Aspekte aufzuzeigen und fasslich zu machen, stellt den Schwerpunkt dieser Handreichung dar. Dabei ist das Ziel dieses Papiers, Empfehlungen für die praktische Berücksichtigung konkreter ethischer Fragen und Aspekte bei der Planung und Entwicklung von KI-Systemen in der Bildung aufzuführen. Die Empfehlungen basieren maßgeblich auf Äußerungen der Teilnehmenden eines Expert_innen-Workshops, der zu diesem Zweck im Kontext des INVITE-Wettbewerbs durchgeführt wurde.

1.2 Expert_innen-Workshop

Im Rahmen des BMBF-geförderten Innovationswettbewerbs INVITE (Digitale Plattform berufliche Weiterbildung) nutzt das Metavorhaben „INVITE-Meta“ das Format der Expert_innen-Workshops zur Diskussion relevanter und aktueller Fragestellungen. Die geplante KI-Verordnung der Europäischen Kommission und die Frage danach, wie auch unabhängig von der Verordnung ethische Aspekte in der Entwicklung KI-gestützter Bildungstechnologien berücksichtigt werden können, sind wichtige Themenfelder, die im Workshop adressiert werden sollten. Der zweistündige Workshop wurde gemeinsam von INVITE-Meta (mmb Institut und DFKI) und der Digitalbegleitung (VDI/VDE-IT) des INVITE-Wettbewerbs organisiert und fand im Oktober 2022 statt. Insgesamt nahmen sieben Expert_innen an dem virtuellen Austausch teil. Die Teilnehmenden sind Vertreter_innen verschiedener relevanter Professionen und Disziplinen, z. B. Bildungswissenschaften, Algorithmenethik, Recht und KI-Entwicklung.

Ziel des Workshops war die gemeinsame Diskussion der durch die EU entworfenen Vorgaben in der KI-Verordnung sowie darauf aufbauend die Erarbeitung von Handlungsempfehlungen für die INVITE-Projekte anhand konkreter Anwendungsfälle aus dem Bereich der beruflichen Weiterbildung. Die Anwendungsfälle und Handlungsempfehlungen werden im weiteren Verlauf dieser Handreichung vorgestellt.

2 KI-gestützte Bildungstechnologien – Drei Anwendungsfälle basierend auf INVITE-Entwicklungen und Expert_innenfeedback

Die Diskussion relevanter Kriterien zur Prüfung der Unbedenklichkeit von KI-Systemen erfolgte im Rahmen des Expert_innen-Workshops anhand dreier (fiktiver) Anwendungsfälle, die aus INVITE-Projekten abgeleitet wurden und einen Ausschnitt der tatsächlich beforschten Szenarien und somit auch Funktionalität und Zweck verschiedener KI-Systeme im Wettbewerb darstellen. Die folgende Vorstellung der diskutierten Anwendungsfälle ist jeweils durch das direkte Feedback und konkrete fallbezogene Empfehlungen der Expert_innen ergänzt. Die übergreifende Sammlung relevanter sozio-ethischer Aspekte und Empfehlungen der Expertenrunde sowie eine Zusammenstellung empfehlenswerter Anknüpfungspunkte und Ressourcen folgen in den Kapiteln 3 und 4.

2.1 Anwendungsfall 1: Auswahl passender Weiterbildungsangebote

Kurzbeschreibung: Die Mitarbeiterin des Versicherungsunternehmens *Betasure* möchte sich im Bereich Risikomanagement weiterbilden. Die verschiedenen Weiterbildungsangebote, die sie zum Thema im Internet findet, sind in ihrer inhaltlichen Ausrichtung sehr verschieden und begründen sehr unterschiedliche Karriereperspektiven. Sie meldet sich bei einer Weiterbildungsplattform eines renommierten Bildungsanbieters an und lässt sich vom System eine passende Weiterbildung mit anerkanntem Abschluss empfehlen.

Zweck des Systems: Anhand eines kurzen digitalen Assessments ihres Wissensstands zum Thema und ihrer Karriereinteressen, erhält die Mitarbeiterin vom KI-System individuelle Weiterbildungsvorschläge verbunden mit einer entsprechenden Buchungsfunktion. Dabei werden nur solche Weiterbildungen vorgeschlagen, für die alle Zugangsvoraussetzungen erfüllt sind. Diese Einschätzung erfolgt automatisiert durch das KI-System auf Basis der hochgeladenen Abschluss- und Arbeitszeugnisse der Mitarbeiterin.

Verwendete Datengrundlage: Abschluss- und Arbeitszeugnisse, Kurzassessments des Wissensstands, Interessensabfrage

Feedback der Expertenrunde zu Anwendungsfall 1:

- Der Anwendungsfall wirft die Frage auf, ob den Nutzenden hier die Handlungsfreiheit eingeschränkt wird. Dabei sollte die User Agency (Handlungsmöglichkeiten der Nutzenden) stets im Vordergrund stehen. Zuerst sollte betont werden, wozu ein System eingesetzt wird, z. B. könnten erst die Interessen erfragt werden und dann die Qualifikationsanforderungen erhoben werden.
- Es sollten Alternativoptionen aufgezeigt werden.
- Zur Qualität der Ergebnisse: Lebensläufe und Arbeitszeugnisse sind bislang meist wenig standardisiert, die Textanalyse dürfte in diesen Fällen eher schlecht funktionieren.

2.2 Anwendungsfall 2: Begleitung des Weiterbildungsprozesses

Kurzbeschreibung: Der Mitarbeiter des Energietechnikunternehmens *GammaHeat* absolviert aktuell eine dreimonatige digitale Weiterbildung eines zertifizierten Bildungsanbieters zum Thema Wärmepumpen. Die Weiterbildung besteht aus passgenau ausgewählten Modulen, die im Verlauf der Weiterbildung adaptiv an Wissensstand, Motivation und zeitliche Verfügbarkeit des Lernenden angepasst werden. Zudem wird der Lernprozess durch einen sogenannten „Learning Companion“ begleitet. Dabei handelt es sich um einen virtuellen Lernbegleiter, der mit dem Lernenden kommuniziert und an ausgewählten Aufgaben spielerisch zusammenarbeitet.

Zweck des Systems: Der KI-basierte Lernbegleiter erkennt frühzeitig Verständnis- und Wissenslücken und passt die präsentierten Lernmaterialien entsprechend automatisch an. Der virtuelle Lernbegleiter kann zudem inhaltliche und administrative Fragen des Lernenden beantworten und motiviert durch die kollaborativen und spielerischen Elemente die Auseinandersetzung mit den Fachinhalten.

Verwendete Datengrundlage: Lernaktivitätsdaten, Chateingaben des Lernenden, Quizergebnisse des Lernenden

Feedback der Expertenrunde zu Anwendungsfall 2:

- Allgemeine Bedenken bestehen gegenüber dem Einsatz von Chatbots bei komplexen Aufgaben wie Lernunterstützung (im Gegensatz zu linearen Dienstleistungen wie Bestellungen o. Ä.). Die Wahrscheinlichkeit, dass Frustration mit dem Chatbot aufkommt, ist beim Lernen sehr hoch. Fragwürdig ist also, ob der Lernbegleiter als Ziel des technischen Systems tatsächlich funktioniert. Man sollte hier eher menschliche Unterstützung nutzen, um komplexe Unterstützung zu bieten.
- **User Agency/Autonomie:** Daher sollte die Möglichkeit des „Opt-out“ (oder andersherum ein „Opt-in“) bei dem Chatbot gegeben sein.

2.3 Anwendungsfall 3: Einsicht in die Lernaktivitäten der Beschäftigten

Kurzbeschreibung: Das Kunststoffunternehmen *Deltaplast* nutzt eine neuartige Learning-Experience-Plattform für die Weiterbildung der eigenen Belegschaft. Auf der Plattform können sich die einzelnen Mitarbeiter_innen, die Personalabteilung, aber auch externe Weiterbildungsanbieterinnen rollenbasiert anmelden und verschiedene Funktionen nutzen.

Zweck des Systems: Die Personalabteilung erhält über Dashboards Einsicht in die Lernaktivitäten der Mitarbeiter_innen. Anhand der Lernaktivitäten, der aktuellen Aufgabengebiete und des Karrierewegs führt das KI-System eine Qualifikationsbedarfsanalyse durch und erstellt individualisierte Vorschläge für die gezielte Weiterbildung einzelner Mitarbeiter_innen. Ein „Predictive Analytics“-Modul ermöglicht zudem die Erstellung von Prognosen zur weiteren Entwicklung einzelner Mitarbeiter_innen, aber auch der Belegschaft einer Abteilung insgesamt.

Verwendete Datengrundlage: Lernaktivitätsdaten der Mitarbeiter_innen, Karriereverläufe im Unternehmen, Lebensläufe der Mitarbeiter_innen, Abgleich mit bundesweiten Trend-Analysen

Feedback der Expertenrunde zu Anwendungsfall 3:

- Data Literacy derjenigen, die es nutzen, ist hier besonders zentral: Können beispielsweise die Mitarbeitenden der Personalabteilung einschätzen, wie aussagekräftig die Ergebnisse und Prognosen des Systems sind?
- Die Anwendung nutzt Learning und Predictive Analytics, das muss in der konkreten Umsetzung besonders kritisch geprüft werden.
- Es ist fragwürdig, ob das System gut funktioniert, denn meist gibt es keine ausreichenden Daten und das System ist instabil gegenüber Veränderungen.
- Transparenz ist hier besonders zentral. Empfehlung: Konzept der „Counterfactual Fairness“ nutzen. Es zeigt auf, wie alternative Antworten zu alternativen Ergebnissen geführt hätten.
- Unklar am System: Trifft das System die Entscheidung über den weiteren Verlauf für Mitarbeiter_innen oder werden lediglich Empfehlungen ausgesprochen? Das ist ein relevanter Unterschied.
- Kritisch: Die Anwendung kann für Überwachung zweckentfremdet werden.

3 Übergreifende Empfehlungen

Die vorgenannten Anwendungsfälle des Einsatzes von KI-Systemen sind nach Einschätzung unserer Expertinnen und Experten in Teilen problematisch, auch wenn sie jeweils einen gewissen Mehrwert zum Status quo der digitalen beruflichen Weiterbildung darstellen. Die Anwendungsfälle wären nach dem derzeitigen Kompromissvorschlag der KI-Verordnung als „hochriskant“ zu klassifizieren (Stand Oktober 2022) und müssten die Mindestanforderungen (Art. 9–15) entsprechend der Verordnung erfüllen. Losgelöst von der konkreten Rechtslage bezogen auf die KI-Verordnung hat die Expert_innengruppe übergreifende Empfehlungen zu den Anwendungsfällen ausgesprochen, die auch auf anders gelagerte Anwendungsfälle der INVITE-Projekte bezogen werden können und hier dargestellt werden:

Betonung des Empfehlungs-Charakters	▪ In allen Fällen sollte für die Nutzenden deutlich ersichtlich werden, dass es sich lediglich um Empfehlungen handelt, die das System produziert. Diese Empfehlungen können auch fehlerhaft sein und stellen keine allgemeinen Wahrheiten oder die tatsächlich bestmögliche Option dar. ▪ Andere Optionen sollten weiterhin sichtbar sein, z. B. indem lediglich eine Rangpriorisierung vorgenommen wird.
Gewährleistung von Transparenz⁴	▪ Die Funktionsweise des KI-Systems sollte den Lernenden am konkreten Beispiel deutlich gemacht werden. ▪ Bereitgestellte technische Dokumentationen sollten die Lernenden in die Lage versetzen, das System einzusehen.
Autonomie der Lernenden	▪ Die Autonomie der Lernenden ist ein zentraler Aspekt, der für verschiedene Systeme kritisch hinterfragt und bestmöglich gefördert werden sollte. ▪ Die Freiheit sich anders zu entscheiden, also der KI-Empfehlung nicht zu folgen, sollte immer gegeben sein. ▪ Als Kernelement selbstregulierten Lernens sollten bspw. Lernziele durch die Lernenden selbst gesteckt werden.
Kognitive Voraussetzungen der Lernenden	▪ Lernende haben gerade in der beruflichen Weiterbildung sehr heterogene Lernvoraussetzungen, welche sich unterschiedlich auf die Nutzung und Akzeptanz von Lerntechnologien auswirken. ▪ Es sollte bei jedem System kritisch geprüft werden, ob Lernende genügend Zeit und kognitive Kapazität haben, um die Entscheidungen oder Vorschläge eines KI-Systems auch zu verarbeiten und zu beurteilen.

⁴ Das Konzept der Erklärbaren KI ist hier ein zentrales und potenziell hilfreiches Konzept: Unter dem Stichwort „explainable artificial intelligence (XAI)“, dt. Erklärbare KI, wird zu Verfahren geforscht, die Nutzer_innen transparent machen, wie Entscheidungen von KI-Systemen zustande kommen. Denkbare Verfahren nutzen z. B. kontrafaktische oder kontrastbildende Erklärungen. Ein Framework für die Anwendung von XAI stellen z. B. Wang et al. 2019 vor. Eine interessante Diskussion der Frage, ob Erklärbarkeit eine rechtliche Anforderung der DSGVO ist, findet sich bei Oshnooei 2022.

Mehrwert eines KI-Systems vs. Selbstzweck	▪ Eine von der Expert_innengruppe kritisch bewertete Beobachtung ist die aktuelle Welle an Ausschreibungen zum Einsatz von KI in verschiedensten Lebensbereichen. Die Notwendigkeit eines KI-Systems erscheint häufig als gegeben und es erfüllt somit lediglich einen Selbstzweck. ▪ Es sollte stets kritisch hinterfragt werden, welchem Zweck der Einsatz von KI dient und welchen konkreten Mehrwert ein KI-System leistet⁵.
Abwägen von Nutzen und Defiziten des KI-Systems	▪ Laut Einschätzung der Expert_innen haben die drei beschriebenen Anwendungen das Potenzial, den Lernprozess der Lernenden zu unterstützen und eine tatsächliche Erleichterung zu verschaffen (z. B. Übersicht im „Weiterbildungsdschungel“ zu bieten). Es ist möglich, dass der Nutzen (z. B. Übersicht) gegenüber etwaigen Defiziten (z. B. Ausblenden von Angeboten) überwiegt. Diese Abwägung sollte durch die Lernenden getroffen werden und ihnen keine unbewussten Nachteile bereiten.
Dynamische Beurteilung von KI-Systemen	▪ Die formale Klassifikation einer KI-Anwendung durch eine externe Prüfung wird das kritische Abwägen verschiedener Faktoren nicht ersetzen. Ein Ergebnis im Sinne von „richtig“ oder „falsch“ wird es in solch einer Betrachtung kaum geben. Die zugrundeliegenden Werte, die zur Einschätzung von positiven oder negativen Konsequenzen des Systems führen, sind gesellschaftlich und kulturell konnotiert und können sich über die Zeit ändern. ▪ Zudem können nicht nur solche Systeme, die im Sinne der KI-Verordnung als hochriskant zu klassifizieren wären, ethisch bedenklich sein. Auch KI-Systeme, die im Sinne der KI-Verordnung ein geringes Risiko tragen, können ethisch höchst bedenklich⁶ sein und sollten nicht voreilig und endgültig als unbedenklich eingestuft werden.

⁵ Weiterführende Literatur hierzu u. a. Raji et al. 2022.

⁶ Ein solches Beispiel bietet der Arbeitsmarktservice-Algorithmus (AMS) in Österreich. Sein Zweck ist es, eine Vorhersage für Arbeitssuchende zu erstellen, darüber wie hoch die Chance auf Wiedereinstieg in den Arbeitsmarkt ist innerhalb eines gewissen Zeitraums. Das System erstellt erste Einschätzungen, basierend auf Statistiken vergangener Jahre, die menschliches Entscheiden und Beraten unterstützen sollen. Die dabei mögliche Stigmatisierung und andere denkbare gesellschaftliche Konsequenzen sind sehr bedenklich (vgl. Österreichische Akademie der Wissenschaften Institut für Technikfolgen-Abschätzung 2023; Lopez 2019).

Einfluss und Bewusstmachung des kulturellen Werteverständnisses	▪ Zu berücksichtigen ist, dass die Betrachtung, ob etwas ethisch tragbar ist, vor dem Hintergrund des eigenen kulturellen Werte- und Normenverständnisses getroffen wird. Daher kann bei Beurteilungen häufig ein westlich-eurozentrierter Bias entstehen. Einen solchen kulturellen Bias kann man durch eine entsprechend diverse, multikulturelle Besetzung von Gremien oder auch Entwicklerteams reduzieren. ▪ Ein ggf. vorhandener Glaubenssatz in Bezug auf Technologieeinsatz sollte wie folgt überschrieben werden: Technische Systeme sind nicht universell.
Bildungsgerechtigkeit als Leitprinzip	▪ Teilhabe und Bildungsgerechtigkeit werden beim Einsatz von KI-Bildungstechnologien bislang wenig thematisiert. In vielen Fällen betreffen Einsatzszenarien fortgeschrittene Lernendengruppen (z. B. an Hochschulen) oder isolierte Karrierewege (z. B. in der beruflichen Weiterbildung). ▪ Die Inklusion bzw. die Ermöglichung von (Weiter-)Bildung für marginalisierte Gruppen, z. B. mit Lernschwierigkeiten, mit Sinnesbehinderungen oder verschiedenen kulturellen Hintergründen, sollte durch technische Systeme verbessert und nicht ausgeblendet werden.
Definition von KI	▪ Hauptproblem bei der gemeinsamen Diskussion über ethische Aspekte von KI-Systemen ist weiterhin die unklare Definition von KI. ▪ Im aktuellen Entwurf der KI-Verordnung findet sich eine sehr breite Definition davon, was KI ist. Viele verschiedene Systeme (beispielsweise auch solche, die lediglich einfache statistische Modellierungen vornehmen) fallen darunter. Während dies einige Entwickler ärgert und vor eine große Menge an Mehrarbeit stellt, sollte dabei die Chance auf eine kritische Auseinandersetzung mit der Breite der vorhandenen Systeme erkannt werden.

4 Weiterführende Ressourcen zur Umsetzung ethischer und vertrauenswürdiger KI-gestützter Bildungstechnologien

Die Expert_innen stellten heraus, dass es bereits viele Richtlinien, Empfehlungen, Standards, etc. zur Implementierung eines ethischen KI-Systems gibt, die genutzt werden können und sollten, um das „Rad nicht neu zu erfinden“. Diese Ressourcen sollten allerdings nicht in ihrer Gesamtheit implementiert werden, sondern als Ausgang zur Erstellung passgenauer Richtlinien, evtl. auch Checklisten, eines individuellen ethischen KI-Systems dienen. Zudem wiesen die Expert_innen darauf hin, dass die Umsetzung von ethischen und rechtlichen Anforderungen sehr komplex, umfassend und zeitaufwändig ist. Gerade im Kontext einer kompletten Förderlinie könnte es sich daher anbieten, dass die Projekte Cluster bilden, um einzelne Aspekte vertieft zu erarbeiten und sich gegenseitig auszutauschen.

Im Folgenden werden die im Workshop herausgestellten Aspekte mit passenden von den Expert_innen genannten sowie recherchierten Ressourcen nach Themen gegliedert und ergänzt.

4.1 Ethisches Handeln im Projekt

Bei der Umsetzung ethischer KI-Systeme sollten ethische Fragen von vornherein mitgedacht werden, so dass die Voraussetzungen für ein ethisches Handeln und Arbeiten im Projekt gegeben sind, welches sich dann wiederum auf die Implementierung eines ethischen KI-Systems positiv auswirken könnte. So könnten z. B. interdisziplinäre Teams eine Perspektivenvielfalt in das Projekt einbringen und somit die ethische Produktentwicklung fördern. Denkbar wäre auch, dass komplementäre Kompetenzen des interdisziplinären Teams eingesetzt werden, z. B. könnten die Kompetenzen eines Technikers um sozio-ethische Kompetenzen ergänzt werden. Auch könnte bei der Umsetzung von Projekten ein prozedurales Denken im Projektverlauf helfen, die Diskussion um die Erfüllung der verschiedenen Anforderungen an ein ethisches KI-System zu strukturieren und im Prozess immer wieder zu überprüfen und zu hinterfragen.

Ressource	Beschreibung	Weiterführende Information
KI-Lebenszyklus	Diese Betrachtung der Entwicklung eines KI-Systems als Lebenszyklus mit z. T. iterativen Phasen ist z. B. den ISO-Standards der Gruppe ISO/IEC JTC 1/SC 42 gemein. Hierunter finden sich Standards zu den Prozessen eines KI-Lebenszyklus (ISO/IEC DIS 5338) sowie ein Framework für den Lebenszyklus von Daten (ISO/IEC DIS 8183). Auch eine AIGO Arbeitsgruppe entwickelte ein Lebenszyklusmodell eines KI-Systems. Dieses Modell unterscheidet die folgenden vier Phasen bei der Entwicklung eines KI-Systems: 1. Design, Daten und Modellierung 2. Verifizierung und Validierung 3. Einführung 4. Betrieb und Monitoring	OECD (2023). Künstliche Intelligenz in der Gesellschaft. Online abrufbar unter: https://www.oecd-ilibrary.org/sites/64ae49f7-de/index.html?itemId=/content/component/64ae49f7-de#section-d1e981 (zuletzt abgerufen am 03.01.2023). ISO (2023). Standards by ISO/IEC JTC 1/SC 42 Artificial intelligence.

	Ethische Überlegungen können in allen vier Phasen relevant sein.	Online abrufbar unter: https://www.iso.org/committee/6794475/x/catalogue/p/1/u/0/w/0/d/0 (zuletzt abgerufen am 03.01.2023).
MEESTAR	Ethische Überlegungen sollten aus interdisziplinärer Perspektive im laufenden Entwicklungsprozess immer wieder geführt werden. MEESTAR – das Modell zur ethischen Evaluation sozio-technischer Arrangements – kann dabei helfen. Das Modell wurde im Kontext Pflege/altersgerechte Assistenzsysteme entwickelt, ist aber auch gut für Technologien außerhalb des Pflegebereiches einsetzbar. In üblicherweise zweitägigen Workshops werden ethische Herausforderungen und Dimensionen einer Technologie in der Tiefe herausgearbeitet und mögliche Handlungsoptionen für die weitere Projekt- bzw. Technologieentwicklung abgeleitet.	Manzeschke et al. (2013). Ethische Fragen im Bereich Altersgerechter Assistenzsysteme. VDI/VDE: Berlin. Online abrufbar unter: https://www.researchgate.net/publication/256474089_Ethische_Fragen_im_Bereich_altersgerechter_Assistenzsysteme_Ergebnisse_der_Studie (zuletzt abgerufen am 03.01.2023).

4.2 Rechtliche Grundlagen

Aus Gründen der Rechtssicherheit sollten die jeweils geltenden Vorschriften bekannt sein. Maßgebend sind hier (geplante) Gesetze, Richtlinien und Verordnungen rund um KI, Daten und Prinzipien, wie Datenschutz oder Barrierefreiheit. Die nachfolgende Tabelle gibt einen Kurzüberblick relevanter rechtlicher Grundlagen in der Bundesrepublik. Sie erhebt keinen Anspruch auf Vollständigkeit.

Rechtliche Grundlage	Beschreibung	Weiterführende Information
KI-Verordnung	Mit der KI-Verordnung (Artificial Intelligence Act bzw. AI Act) wird ein EU-weit einheitlicher Rechtsrahmen für die Entwicklung, Vermarktung und Verwendung Künstlicher Intelligenz geschaffen. Vorgesehen sind vier Risikoklassen. KI mit einem unannehmbaren Risiko wird verboten. KI mit einem hohen Risiko unterliegt hohen Compliance-Vorgaben. Systeme, die mit einem geringen oder minimalen Risiko assoziiert sind, müssen Transparenzvorgaben erfüllen bzw. unterliegen keinen regulatorischen Vorgaben. Die Verordnung wird voraussichtlich Anfang 2023 verabschiedet und gilt ab dem Jahr 2026 unmittelbar in der gesamten EU.	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=cele_x%3A52021PC0206 „DIE REGULIERUNG VON KÜNSTLICHER INTELLIGENZ. Auswirkungen der geplanten KI-Verordnung auf Bildungsplattformen“ von Sebastian Straub, LL.M., VDI/VDE-IT in der aktuellen Fassung)

Richtlinie über KI-Haftung	Die EU-Richtlinie über KI-Haftung befindet sich zum Zeitpunkt der Erstellung dieser Publikation noch im Gesetzgebungsprozess. Sie ergänzt die KI-Verordnung und legt fest, wie mit Schäden umzugehen ist, die durch KI-Systeme verursacht wurden.	https://ec.europa.eu/info/sites/default/files/1_1_1976_05_prop_dir_ai_en.pdf
Data Governance Act	Der Daten-Governance-Rechtsakt ist am 23.06.2022 in Kraft getreten und wird ab September 2023 anwendbar sein. Er ist Teil der Europäischen Datenstrategie. Ziel der Verordnung ist es, den Austausch von Daten branchenübergreifend und EU-weit zu ermöglichen.	https://eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX%3A52020PC0767
Data Act	Das Daten-Gesetz befindet sich zum Zeitpunkt der Erstellung dieser Publikation noch im Gesetzgebungsprozess. Es ist Teil der Europäischen Datenstrategie. Der Data Act zielt darauf ab, mehr Daten verfügbar zu machen. In der Entwurfsfassung vom 23.02.22 beschränkt sich der Anwendungsbereich auf physische Produkte, die Daten sammeln und übermitteln. Darunter können u. a. Smart Devices fallen. Im Gegensatz dazu soll Hardware, wie Smartphones oder PCs nicht zum Anwendungsbereich des Gesetzes zählen.	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celext%3A52022PC0068 https://ec.europa.eu/commission/presscorner/detail/de/ip_22_1113
Digital Services Act	Der Digital Services Act (Gesetz über digitale Dienste) ist am 16.11.2022 in Kraft getreten. Anwendbarkeit erlangt die EU-Verordnung am 17.02.2024. Ziel des Gesetzes ist es u. a. einen Transparenz- und Rechenschaftsrahmen für Online-Plattformen zu schaffen. So gibt beispielsweise Artikel 27 vor, dass Online-Plattformen, die Empfehlungssysteme verwenden, im Rahmen der Allgemeinen Geschäftsbedingungen über die wichtigsten zugrundeliegenden Parameter informieren und Nutzer_innen aufklären, wie diese geändert bzw. beeinflusst werden können.	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065&qid=1667980362928 https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_de

Datenschutz-Grundverordnung (DSGVO)	Die Europäische Datenschutz-Grundverordnung ist im Mai 2018 in Kraft getreten. Sie schützt natürliche Personen, wenn bestimmte Daten über sie verarbeitet werden. Diese Daten werden als personenbezogene Daten bezeichnet. Es handelt sich dabei um alle Informationen, die eine natürliche Person identifizieren (z. B. Name) oder identifizierbar machen (z. B. Kennnummer, die einem Namen zugeordnet werden kann). Die Verordnung formuliert den rechtlichen Rahmen für die Verarbeitung dieser personenbezogenen Daten.	https://www.datenschutz-grundverordnung.eu/
Bundesdatenschutzgesetz	Das Bundesdatenschutzgesetz ergänzt seit Mai 2018 die Datenschutz-Grundverordnung (DSGVO) um die Bereiche, die von der DSGVO nicht spezifisch geregelt sind und der nationalen Gesetzgebung unterliegen. Ergänzende Bestimmungen umfassen beispielsweise die Verarbeitung besonderer personenbezogener Daten (z. B. Daten von Beschäftigten).	http://www.gesetze-im-internet.de/bdsg_2018/
Gesetz zur Gleichstellung von Menschen mit Behinderungen (BGG)	Das Gesetz zur Gleichstellung von Menschen mit Behinderungen adressiert die gleichberechtigte Teilhabe von Menschen mit Behinderung am gesellschaftlichen Leben. Das Gesetz richtet sich an Träger der öffentlichen Gewalt und enthält u. a. Regelungen zur Barrierefreiheit von Software.	http://www.gesetze-im-internet.de/bgg/ https://www.bundesfachstelle-barrierefreiheit.de/DE/Fachwissen/Informationstechnik/Barrierefreie-Software/barrierefreie-software_node.html

4.3 KI-Normen und -Standards

KI-Normen und -Standards haben das Ziel einen klaren Handlungsrahmen für KI-Systeme zu schaffen. Für diverse Formen und Entwicklungsprozesse von KI existieren bereits Normen und Standards. Weitere befinden sich in der Entwicklung. Gleichwohl existieren noch keine explizit für den Bildungsbereich entwickelten KI-Standards⁷.

Eine Übersicht von Normen für KI in der Arbeitswelt findet sich z. B. bei Becker et al. (Becker et al. 2021) (siehe insbesondere Übersicht auf Seite 8). Auch die Deutsche Normungsroadmap Künstliche Intelligenz bietet eine Übersicht (Wahlster und Winterhalter 2020). Diese Publikation beschreibt außerdem Normungs- und Standardisierungsbedarfe für KI unter ethischen Gesichtspunkten. Darüber hinaus arbeitet der weltweite Berufsverband IEEE (Institute of Electrical and Electronics Engineers) an verschiedenen Standards im Zusammenhang mit KI-Systemen. Das Gremium IEEE Standards Association (IEEE SA) entwickelt in der IEEE P7000™ Reihe ethische Grundsätze und Standards von autonomen und intelligenten Systemen. Davon erscheinen insbesondere zwei Standards für den Kontext der (beruflichen Weiter-) Bildung hilfreich.

KI-Norm bzw. -Standard	Beschreibung	Weiterführende Information
IEEE 7000-2021 IEEE Standard Model Process for Addressing Ethical Concerns during System Design	Dieser Standard legt eine Reihe von Prozessen fest, mit denen Organisationen ethische Werte in allen Phasen der Konzeption und -entwicklung einbeziehen, priorisieren und kommunizieren können.	https://standards.ieee.org/ieee/7000/6781/
IEEE 7001-2021 IEEE Standard for Transparency of Autonomous Systems	Dieser Standard legt messbare und überprüfbare Transparenzniveaus für physische und nicht-physische autonome Systeme fest, die direkt physischen, psychischen, gesellschaftlichen, wirtschaftlichen, ökologischen oder rufschädigenden Schaden verursachen können. Anwendungsbeispiele könnten Roboter für betreutes Wohnen oder medizinische Diagnose-(Empfehlungs-)Systeme oder Chatbots sein.	https://standards.ieee.org/ieee/7001/6929/

⁷ Weitere Hinweise zu KI-Standards finden Sie auch in einem weiteren Dossier aus dem Kontext des INVITE-Meta-Projekts, welches im Frühjahr 2023 erscheinen wird (Rashid et al., im Druck).

4.4 Zertifizierungen

Es ist davon auszugehen, dass Zertifizierungen von KI-Systemen im Kontext der KI-Verordnung an Relevanz gewinnen werden. Durch Zertifizierungen, die auf bestehenden Normen aufsetzen, werden konkrete prüfbare Aspekte aufgeführt. Im Fall der KI-Zertifizierung könnten diese Aspekte Autonomie, Fairness, Transparenz, Verlässlichkeit, Sicherheit und Datenschutz sein. Die Vorteile von Zertifizierungen liegen z. B. in der Schaffung von Vertrauen bei Nutzer_innen in das KI-System. Allerdings sollte man nicht übersehen, dass Zertifizierungen zeit- und kostenaufwändig sind und dadurch kleinere Unternehmen benachteiligt sein könnten⁸.

Zertifizierung	Beschreibung	Weiterführende Information
KI-Zertifizierung in Nordrhein-Westfalen	Das Papier ist in interdisziplinärer Zusammenarbeit mit Vertreter_innen aus Philosophie, Recht und Informatik entstanden und beschreibt eine erste Grundlage für die Entwicklung einer KI-Zertifizierung in NRW. Ausgehend von sechs Handlungsfeldern soll ein KI-Prüfkatalog erarbeitet werden. Die Handlungsfelder umfassen: 1. Autonomie und Kontrolle 2. Fairness 3. Transparenz 4. Verlässlichkeit 5. Sicherheit 6. Datenschutz	https://www.iais.fraunhofer.de/content/dam/iais/KINRW/Whitepaper_KI-Zertifizierung.pdf

4.5 Leitfäden, Handreichungen und Tools

Es gibt eine Vielzahl an Leitlinien bzw. Handreichungen zur ethischen Entwicklung und Implementierung von KI-Systemen („Ethics-by-design“). Ein Online-Inventar, das von einem Projekt des gemeinnützigen Unternehmens Algorithm Watch erstellt wurde, listet zurzeit 167 recherchierte Leitfäden zur ethischen Entwicklung von algorithmischen Entscheidungssystemen (AW AlgorithmWatch gGmbH 2023). Auch eine Inhaltsanalyse öffentlich zugänglicher Ethik-Leitfäden (Institute for Digital Transformation in Healthcare GmbH 2020) zeigt, wie umfassend sich die Bandbreite an Stakeholdern, die Leitfäden oder andere Formen von Handreichungen hierzu entwickelt haben, darstellt. Im Rahmen dieser Analyse waren die häufigsten in den Leitlinien verankerten Werte: Transparency, Beneficence, Impartiality (Bias⁹), Explainability und Autonomy.

⁸ Für eine gute Übersicht von Argumenten für und gegen eine KI-Zertifizierung siehe Mangelsdorf et al. 2021.

⁹ Als weiterführende Literatur zu Bias in Daten sei an dieser Stelle auf das nachfolgende Paper verwiesen, das drei Formen von Bias in Daten unterscheidet (gesellschaftlicher, technischer und sozio-technischer Bias) (Lopez 2021).

Eine kleine Auswahl an Ethik-Leitfäden kann der nachfolgenden Tabelle entnommen werden.

Leitfaden, Handreichung bzw. Tool	Beschreibung	Weiterführende Information
Ethische Leitlinien für Lehrkräfte über die Nutzung von KI und Daten für Lehr- und Lernzwecke	Im September 2022 erschienen diese Leitlinien der Expert_innengruppe für Künstliche Intelligenz und Daten in der Bildung. Sie richten sich an Lehrkräfte und enthalten u. a. Leitfragen zum Einsatz von KI-Systemen. Außerdem enthält das Dokument Hinweise, wie die Leitfragen in der Praxis eingesetzt werden können und welche Kompetenzen Lehrkräfte benötigen, um KI-Systeme unter Berücksichtigung ethischer Gesichtspunkte zu nutzen.	Europäische Union (2022). Ethische Leitlinien für Lehrkräfte über die Nutzung von KI und Daten für Lehr- und Lernzwecke. Online abrufbar unter: https://op.europa.eu/o/opportal-service/download-handler?identifizier=d81a0d54-5348-11ed-92ed-01aa75ed71a1&format=pdf&language=de&productionSystem=cellar&part= (zuletzt abgerufen am 13.12.2022).
Algorules	Die Algorules umfassen neun Regeln für die Gestaltung algorithmischer Systeme. Für Entwickler_innen gibt es einen Praxisleitfaden, der bei der Anwendung der Regeln unterstützt. Seit 2018 werden die Algorules entwickelt.	https://algorules.org/fileadmin/files/user_upload/Algo.Rules_Praxisleitfaden.pdf
ALTAI (The Assessment List for Trustworthy Artificial Intelligence)	Das ALTAI Portal bietet ein Tool zur Selbsteinschätzung der Vertrauenswürdigkeit eines KI-Systems im Entwicklungsstadium. Das Tool basiert auf dem Konzept der Vertrauenswürdigen KI, das in den Ethikleitlinien für vertrauenswürdige künstliche Intelligenz entwickelt wurde.	European Commission (2022). ALTAI portal. Online abrufbar unter: https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal (zuletzt abgerufen am 13.12.2022). Hochrangige Expertengruppe für künstliche Intelligenz (2019). Ethikleitlinien für vertrauenswürdige künstliche Intelligenz. Online abrufbar unter: https://op.europa.eu/de/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 (zuletzt abgerufen am 13.12.2022).

Algorithmen in der Entscheidungsfindung – Leitfaden zu Verantwortlichkeit und Rechenschaft	Ein Online-Tool zur Bestimmung und Verteilung von Verantwortung bzgl. algorithmischer Entscheidungsfindung. Die Handreichung enthält außerdem Hinweise, worauf bei der Einführung eines algorithmischen Entscheidungssystems zu achten ist.	https://wien.arbeiterkammer.at/interessenvertretung/arbeitsdigital/DataPolitics/Algorithmen_in_der_Entscheidungsfindung.html
CapAI Procedure	Eine Handreichung, die ein Tool enthält, um die Konformität der eigenen KI-Anwendung mit der KI-Verordnung zu überprüfen. Das Vorgehen wird in fünf Schritte unterteilt – entlang des Produktzyklus einer KI-Anwendung: design, development, evaluation, operation, und retirement.	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4064091

4.6 Selbstverpflichtung

Neben den oben angeführten rechtlichen Grundlagen, die verpflichtende Regeln beschreiben (hard governance), gibt es Regeln, die zwar nicht verpflichtend sind, aber dennoch eine regulierende Wirkung entfalten können (soft governance)¹⁰. Soft governance kann die Form von Selbstverpflichtungen der Organisationen, die KI-Systeme entwickeln und/oder einsetzen, annehmen. Um eine Selbstverpflichtung zu erstellen, sollte die Organisation ganz pragmatisch vorgehen. Es sollte untersucht werden, welche Geschäftsleitlinien oder Ethikleitlinien bereits in der Organisation existieren und wie das KI-System sich zu diesen Leitlinien compliant verhalten könnte. Zudem sollten bereits existierende Selbstverpflichtungen als Referenz herangezogen werden, um daraus dann eine maßgeschneiderte Selbstverpflichtung für das ethische KI-System zu entwickeln.

¹⁰ Weitere Informationen zu hard und soft governance in der EU finden sich z. B. in Maggetti 2015.

Selbstverpflichtung	Beschreibung	Weiterführende Information
KI-Ethik: Vom Prinzip zur individuellen Verantwortung	Im Rahmen des Roundtable KI der Gesellschaft der Informatik fanden Workshops zu KI-Selbstverpflichtungen statt. Die daraus entstandenen Papiere vergleichen die ethischen Selbstverpflichtungen der Workshopteilnehmenden und geben u. a. eine Übersicht technischer und sozialer Prinzipien sowie der implementierten Prozesse und Verantwortlichkeiten zur Operationalisierung dieser Prinzipien.	https://roundtable-ki.gi.de/ https://roundtable-ki.gi.de/fileadmin/PR/Roundtable-KI/Inputpapier_RTeKI_2022-09-13.pdf
KI-Ethik-Label (Bertelsmann Stiftung und VDE)	Das entwickelte KI-Ethik-Label ist dem von Haushaltsgeräten bekannten Energieeffizienzlabel nachempfunden und soll für verschiedene Stakeholder eine schnelle Orientierung, Transparenz und Vergleichbarkeit bieten.	https://www.ai-ethics-impact.org/resource/blob/1990244/784c99a910c18247c5b07bc5d62ac2bf/aieig---report---download-hb---de-data.pdf https://www.reframetech.de/2020/04/02/f

5 Literaturverzeichnis

- AW AlgorithmWatch gGmbH (2023): AI Ethics Guidelines Global Inventory. Berlin. Online verfügbar unter <https://inventory.algorithmwatch.org/>, zuletzt aktualisiert am 04.01.2023, zuletzt geprüft am 04.01.2023.
- Becker, Nikolas; Junginger, Pauline; Martinez, Lukas; Krupka, Daniel (2021): KI in der Arbeitswelt. Übersicht einschlägiger Normen und Standards. In: *ExamAI – KI Testing & Auditing*. Online verfügbar unter https://testing-ai.gi.de/fileadmin/PR/Testing-AI/ExamAI_Normenuebersicht_2021_04.pdf, zuletzt geprüft am 03.01.2023.
- 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. Online verfügbar unter https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0019.02/DOC_1&format=PDF, zuletzt aktualisiert am 2021, zuletzt geprüft am 16.01.2023.
- Institute for Digital Transformation in Healthcare GmbH (2020): Europaweite Leitlinien zur digitalen Ethik im Vergleich. Online verfügbar unter https://www.transforming-healthcare.com/wp-content/uploads/2020/06/idigit2020Europaweite_Leitlinien_zur_digitalen_Ethik_im_Vergleich_executive_summary.pdf, zuletzt geprüft am 03.01.2023.
- Lopez, Paola (2019): Reinforcing Intersectional Inequality via the AMS Algorithm in Austria. Graz (Proceedings of the 18th Annual STS Conference Graz 2019). Online verfügbar unter <https://digilib.tugraz.at/download.php?id=5e29a88e0e34f&location=browse>, zuletzt aktualisiert am 2019, zuletzt geprüft am 27.01.2023.
- Lopez, Paola (2021): Bias does not equal bias: a socio-technical typology of bias in data-based algorithmic systems. In: *Internet Policy Review* 10 (4). DOI: 10.14763/2021.4.1598.
- Maggetti, Martino (2015): Hard and Soft Governance. In: Kennet Lynggaard (Hg.): Research methods in European Union studies. 1. publ. Basingstoke: Palgrave Macmillan (Palgrave studies in European Union politics), S. 252–265.
- Mangelsdorf et al. (2021): Die Zertifizierung von KI: Mehr Sicherheit für alle – oder unnötiger Ballast? In: *iit perspektive* (58). Online verfügbar unter https://www.iit-berlin.de/wp-content/uploads/2021/04/2021_04_30_iit-perspektive_Nr-58_Zertifizierung_von_KI.pdf, zuletzt geprüft am 03.01.2023.
- Oshnooei, Nima (2022): Data Science: Künstliche Intelligenz (KI) & das Transparenzgebot. Hg. v. intersoft consulting services AG. Online verfügbar unter <https://www.dr-datenschutz.de/data-science-kuenstliche-intelligenz-ki-das-transparenzgebot/>, zuletzt aktualisiert am 28.07.2022, zuletzt geprüft am 03.01.2023.
- Österreichische Akademie der Wissenschaften Institut für Technikfolgen-Abschätzung (2023): Der AMS-Algorithmus. Online verfügbar unter <https://www.oeaw.ac.at/ita/projekte/der-ams-algorithmus>, zuletzt aktualisiert am 03.01.2023, zuletzt geprüft am 03.01.2023.
- Raji, Inioluwa Deborah; Kumar, I. Elizabeth; Horowitz, Aaron; Selbst, Andrew (2022): The Fallacy of AI Functionality. In: FAccT 2022. 2022 5th ACM Conference on Fairness, Accountability, and Transparency : June 21-24, 2022, Seoul, South Korea. FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency. Seoul Republic of Korea, 21 06 2022 24 06 2022. New York, New York: The Association for Computing Machinery (ICPS), S. 959–972. Online verfügbar unter

- https://facctconference.org/static/pdfs_2022/facct22-78.pdf, zuletzt geprüft am 03.01.2023.
- Straub, Sebastian (2022): Die Regulierung von Künstlicher Intelligenz. Auswirkungen der geplanten KI-Verordnung auf Bildungsplattformen. Berlin. Online verfügbar unter https://www.ueberaus.de/l.php?files%7Cinvite%40ueberaus.de%7C%2F05_Handreicungen%2F%C3%9Cberblick_KI-Verordnung.pdf%7C/ (INVITE-internes Dokument).
- Wahlster, W.; Winterhalter, C. (Hg.) (2020): Deutsche Normungsroadmap Künstliche Intelligenz. Online verfügbar unter <https://www.din.de/resource/blob/772438/6b5ac6680543eff9fe372603514be3e6/normungsroadmap-ki-data.pdf>, zuletzt geprüft am 03.01.2023.
- Wang, Danding; Yang, Qian; Abdul, Ashraf; Lim, Brian Y. (2019): Designing Theory-Driven User-Centric Explainable AI. In: *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Artikel 601, S. 1–15. Online verfügbar unter <https://dl.acm.org/doi/10.1145/3290605.3300831>, zuletzt geprüft am 03.01.2023.