Themenkartenstapel:

Künstliche Intelligenz in der Arbeitswelt – Was ist humanzentrierte KI?

Themenkartenstapel Überblick

Themenkarten

Was ist humanzentrierte KI? 3

Was bedeutet Humanzentrierung und wie erreicht man sie? 6

Welche Methoden gibt es, um hKI umzusetzen? 8

Welche Systeme wirken in der hKI zusammen? 9

Best-Practice-Beispiele 10

Literatur 11

Was ist humanzentrierte KI?

Die Konzepte, für die humanzentrierte Künstliche Intelligenz (KI) steht, begegnen uns unter verschiedenen Namen und Varianten. Erklärbare KI, ethische KI, algorithmische Transparenz. Alle haben gemeinsam, dass das Zusammenspiel zwischen Mensch und Technologie in das Zentrum der Aufmerksamkeit gesetzt wird. Denn während KI in einer wachsenden Menge von alltäglichen und industriellen Anwendungsfällen eingesetzt wird, haben Forscher:innen verschiedener Disziplinen rechtliche, ethische und soziale Probleme identifiziert. Das Erscheinen von ChatGPT stellte einen Einschnitt in die Geschichte der KI dar, denn seit November 2022 können erstmals auch Menschen ohne technische Kenntnisse in natürlicher Sprache mit einer KI interagieren. Die wissenschaftliche Gemeinschaft sowie viele Initiativen und Komitee haben jedoch schon lange davor angefangen, Richtlinien für die Nutzung von KI zu entwickeln. So gibt es unter anderem Richtlinien von der OECD, der Enquete-Kommission KI des deutschen Bundestags und der Europäischen Kommission.

Während KI von Jahr zu Jahr leistungsfähiger wird, bleiben grundsätzliche Herausforderungen und Risiken. Die Risiken resultieren direkt daraus, dass gewichtige Entscheidungen basierend auf der Ausgabe von Modellen maschinellen Lernens getroffen werden und das Leben von Menschen beeinflussen.

Man kann die Prinzipien von humanzentrierter KI drei Kategorien zuteilen:

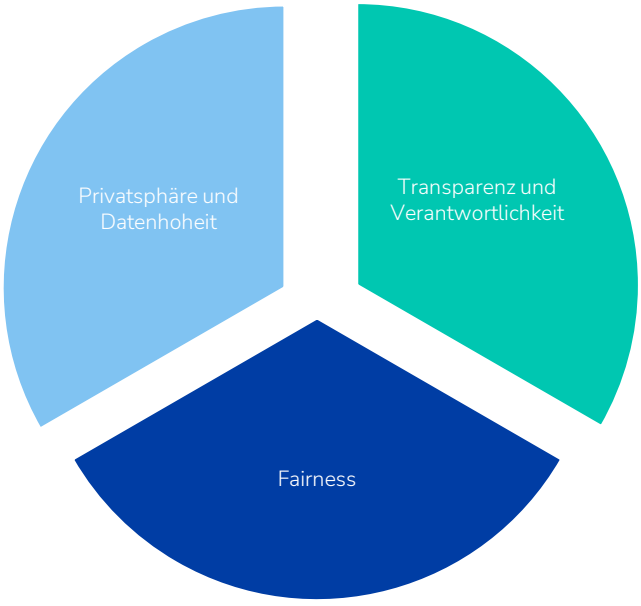

Abbildung 1: eigene Darstellung, adaptiert von Lepri et. al. (2021)

In jeder der drei Kategorien gibt es Risiken, die hier mit je einem Beispiel erklärt werden:

- **Privatsphäre und Datenhoheit:** Sensible Daten über menschliche Charakteristika und Verhaltensweisen können durch KI-Systeme verarbeitet werden und zu Erkenntnissen über private Informationen führen. Absichtlicher oder unabsichtlicher Missbrauch dieser Informationen durch die Nutzer:innen und unbefugte Dritte kann die Folge sein.
    - **Beispiel:** Durch Nutzungsdaten und digitale Fußabdrücke aus sozialen Medien kann auf die psychologischen Eigenschaften einer Person geschlossen werden. Diese

    - können unter anderem genutzt werden, um individuell zugeschnittene Werbung zu schalten.
- Transparenz und Verantwortlichkeit: KI-Entscheidungen werden durch Prozesse erzeugt, die auf komplexen statistischen Methoden basieren und für Menschen nicht intuitiv verständlich sind. Es stellte sich als schwierig heraus, nachzuvollziehen, warum und auf Basis welcher Informationen eine KI zu einer Entscheidung gelangt. Manchmal ist das Absicht, wie wenn eine Firma zu kommerziellen Zwecken eine eigene KI baut und Nachahmer befürchte. Oft jedoch besteht das Problem darin, dass das KI-Modell eine „intrinsische Trübung“ hat, also von Natur aus schwer zu interpretieren ist.
    - Beispiel: Eine KI sortiert am Fließband fehlerhafte Artikel aus. Der Mensch, der regelmäßig die Sortierung prüft, kann den Entscheidungsprozess jedoch schwer nachvollziehen, warum ein bestimmter Artikel ausgemustert wurde: Liegt es am Gewicht, an der Verpackung, oder doch am Etikett?
- Fairness: Ein „Bias“ ist eine systematische Verzerrung von Daten. Wenn die Daten, mit denen eine KI trainiert wird, eine solche systematische Verzerrung inne haben, dann wird die Ausgabe der KI diese falsche Gewichtung reflektieren. Die Konsequenz können Diskriminierung und blinde Flecken in der Entscheidungsfindung sein.
    - Beispiel: Die Fotodatenbank, mit der eine KI auf das Erkennen von Vögeln trainiert wurde, enthält nur Fotos von in Zentraleuropa heimischen Vögeln. Tukane und Pelikane werden von der KI daher viel seltener korrekt identifiziert als Spechte und Finken.

Die humanzentrierte KI ist also ein soziales und technologisches Grundgerüst, welches um eine Reihe zentraler Prinzipien herum konstruiert wird, um die Fähigkeiten von Menschen zu erweitern und das menschliche Wohlbefinden dabei zu erweitern. Der Einsatz von hKI soll positive ökologische und soziale Effekte aufweisen. Die Mindestanforderung ist also, dass ökologische und soziale Faktoren mit dem Einsatz von KI nicht schlechter gestellt sind als ohne KI.
An hKI arbeiten deshalb Experten verschiedener Disziplinen mit, die in den verschiedenen Schritten der Einführung und des alltäglichen Einsatzes von KI mitarbeiten. Neben technischen Kenntnissen, wie die obigen Kategorien in Nullen und Einsen übersetzt werden können, braucht es die Hilfe von Psycholog:innen und
Die folgenden Themenkarten erklären leicht verständlich die beiden zentralen Konzepte der KI und der Humanzentrierung, stellen die wichtigsten Methoden vor, und zeigt anhand von Anwendungsbeispielen aus Medizin, Bildung und Industrie, wie humanzentrierte KI dazu beiträgt, das Mensch und KI gemeinsam Probleme bearbeiten können.

Künstliche Intelligenz: Warum der Vergleich zur menschlichen Intelligenz hinkt

Schon die menschliche Intelligenz ist unheimlich schwer zu definieren, weshalb die Aufgabe Psycholog*innen, Neurowissenschaftler*innen und Biolog*innen bereits seit Jahrzehnten beschäftigt. Der Begriff „Künstliche Intelligenz“, kurz **KI, stünde strikt für die Imitation dieser menschlichen Intelligenz durch Computer**. Da wir diese aber noch nicht vollständig verstanden haben, bringt uns diese Definition für den Alltag nicht weit. Außerdem sind Computer sehr gut darin, einige für Menschen schwierige Probleme fehlerfrei zu lösen (wie z.B. Rechenaufgaben), andere **für Menschen triviale Aufgaben wie Katzen von Hunden zu unterscheiden sind für Computer ungemein schwierig**.

Parallel damit, dass Computer immer mehr komplexe, bisher nur vom Menschen lösbare Probleme bearbeiten können, entwickelt sich auch unsere Vorstellung davon, was eine KI eigentlich ist, weiter. Für lange Zeit wurde das Schachspielen als ultimativer Ausdruck menschlicher Intelligenz empfunden. Heute, über 20 Jahre nachdem ein Computer den besten Schachspieler der Welt schlug, wird die Leistung als simple Kalkulation eines mathematischen Problems gesehen, und nicht als Ausdruck „echter“ Intelligenz gesehen. KI nennen viele nur dann so, wenn sie uns verblüfft. Es folgt, dass häufig die neuesten Technologien aus dem Gebiet der Informatik mit Künstlicher Intelligenz gleichgesetzt werden. Beispielsweise: Große Sprachmodelle á la ChatGPT wurden vor ihrer breiten Anwendung seit Ende 2022 regelmäßig als der Meilenstein festgelegt, an dem der Computer dem menschlichen Intellekt ebenbürtig sei. Heute, Jahre danach, hat die Allgemeinheit ein Verständnis für die technischen Grundlagen dieser „Large Language Models“ erhalten. Grenzen und Schwächen der Technologie sind bekannter, fast jede*r kennt Beispiele, bei denen KI-Modelle offensichtliche Fehler machen oder halluzinieren.

Zusammengefasst: Der Begriff führt also auf die falsche Fährte, indem er Computer an menschlichen Fähigkeiten misst**. Computer sind keine künstlichen Menschen**, sondern komplett eigene Entitäten. Einige Forscher machten Vorschläge, sie „Vorhersagen-Maschinen“ zu nennen. Sie können wie Orakel agieren: **Sie geben Vorhersagen (basierend auf ihrer Programmierung und den Daten, mit denen sie gefüttert wurden), unterliegen aber anders als Menschen keiner Psychologie.** Sie handeln ohne Emotionen oder Gewissen, haben keine Persönlichkeit.

Was bedeutet Humanzentrierung und wie erreicht man sie?

Der Begriff "Humanzentrierung" bezieht sich auf eine Herangehensweise, bei der technologische Entwicklungen und Anwendungen primär auf das Wohl des Menschen und seiner Lebensqualität ausgerichtet werden. Im Kontext der Künstlichen Intelligenz bedeutet das, Systeme so zu gestalten, dass sie die Bedürfnisse, Werte und Ziele der Menschen respektieren und unterstützen.
Die **Arbeitsteilung zwischen Menschen und Technik** wird nach den Vorstellungen humanzentrierter KI als erfolgreich angesehen, wenn Menschen in der Arbeitswelt bestätigt und bestärkt werden.
Ein Whitepaper von Huchler et. al. setzt hier vier Cluster an:

1. **Der Schutz des einzelnen Menschen**
    - Sicherheit und Schutz der Gesundheit
    - Schutz der Daten und eine verantwortungsbewusste Erfassung von Leistung
    - Sensibilität gegenüber menschlicher Vielfalt
    - Diskriminierungsfreiheit
2. **Vertrauenswürdigkeit:**
    - Hohe inhaltliche Datenqualität
    - Nachvollziehbare, erklärbare und widerspruchsfreie Systeme
    - Verantwortung, Haftung und Systemvertrauen
3. **Sinnvolle Arbeitsteilung:**
    - Angemessene Teilung, Entlastung und Unterstützung seitens der KI
    - Klare Handlungsträgerschaft und Kontrolle über die Situation seitens der anwendenden Menschen
    - Soziale Anpassungsfähigkeit des KI-Systems
4. **Förderliche Arbeitsbedingungen:**
    - Handlungsräume für Menschen und reichhaltige Arbeit
    - Ermöglichung des wechselseitigen Lernens des Menschen von der Maschine und umgekehrt

Humanzentrierte KI will nicht den Menschen ersetzen, sondern ihn befähigen und seine Fähigkeiten ergänzen. Diese Systeme sollen das Leben verbessern, Sicherheit bieten und Transparenz gewährleisten. Beispielsweise wird zunehmend betont, dass humanzentrierte KI-Systeme so entwickelt werden sollten, dass sie verständlich, verlässlich und vorhersehbar sind. Um dies zu erreichen, müssen Technikerinnen und Entwicklerinnen eng mit Nutzern zusammenarbeiten und die ethischen, sozialen und kulturellen Implikationen der Systeme berücksichtigen. So sollten humanzentrierte KI-Anwendungen auch Mechanismen bieten, die es den Nutzer*innen ermöglichen, sich aktiv in den Entscheidungsprozess einzubringen, indem sie KI-Entscheidungen hinterfragen, korrigieren oder ablehnen können. Die Verantwortung für Entscheidungen bleibt bei den Menschen, und die KI wird als Werkzeug zur Unterstützung und Erweiterung menschlicher Fähigkeiten eingesetzt. Doch wenn die Verantwortung für die Folgen einer KI-Entscheidung bei den Nutzer:innen und somit bei Unternehmen und ihren Beschäftigten liegt, dann fordert die humanzentrierte KI: **Beschäftigte benötigen klare Regeln, ein ausgeprägtes Verständnis für die Möglichkeiten und Grenzen der KI sowie Vertrauen, dass auch die KI immer im Rahmen der Regeln handelt.**

Themenkarte 3

Weiter soll die Arbeitsteilung in der hKI so geschehen, **dass der Mensch die motivierende, qualifizierende und gesundheitsförderliche Arbeit übernimmt**, während die KI möglichst viele der entgegenwirkenden Arbeiten ausübt.

Schließlich ist der Einführungsprozess gesondert zu betonen. Die Abläufe von der Zielsetzung bis über die Planung und Implementierung bis hin zur Evaluation unter Beteiligung der betroffenen Anwender:innen nennt man **Change Management**, im zweiten Themenkartenstapel „Künstliche Intelligenz in der Arbeitswelt – Die psychologische Perspektive“ wird er vertieft behandelt.

Welche Methoden gibt es, um hKI umzusetzen?

Schnittstellen

Die Gestaltung von Schnittstellen in humanzentrierten KI-Systemen ist ein zentraler Aspekt für die Benutzerfreundlichkeit und Akzeptanz der Technologie. Es ist von entscheidender Bedeutung, dass Schnittstellen intuitiv, zugänglich und für alle Nutzergruppen verständlich sind. Gut gestaltete Schnittstellen geben den Nutzer*innen Sicherheit und erleichtern den Umgang mit der Technologie. Dabei ist Transparenz ein Schlüsselfaktor: Schnittstellen sollen den Nutzern die Funktionsweise der KI verständlich machen, indem sie Einblick in die Entscheidungsprozesse gewähren. Beispielsweise können interaktive Dashboards, die visuelle Informationen über die zugrunde liegenden Daten und Algorithmen bieten, das Vertrauen in das System stärken. Erklärbare KI (Explainable AI) ist eine Methode, die sich hier besonders eignet, da sie den Entscheidungsprozess der KI in einer verständlichen Form darstellt und somit ein wichtiges Werkzeug für die Humanzentrierung darstellt.

Visualisierungen

Visualisierungen sind ein effektives Mittel, um die Arbeitsweise und die Ergebnisse eines KI-Systems zu vermitteln. Sie helfen dabei, komplexe Informationen in einer Form darzustellen, die für Menschen nachvollziehbar ist. Verschiedene Visualisierungsformen wie Diagramme, Heatmaps oder Netzwerkanalysen können eingesetzt werden, um Muster, Zusammenhänge oder Auffälligkeiten zu verdeutlichen, die sich aus den Daten ergeben. Gerade in Bereichen wie der medizinischen Diagnostik oder der Finanzanalyse, in denen große Datenmengen und komplizierte Zusammenhänge eine Rolle spielen, ermöglichen Visualisierungen es Fachleuten, die Informationen schnell zu erfassen und Entscheidungen fundiert zu treffen. Visualisierungen tragen zur Transparenz bei, indem sie den Nutzer*innen ermöglichen, die Ergebnisse der KI zu hinterfragen und kritisch zu prüfen. Wichtig ist auch, dass diese Visualisierungen auf unterschiedliche Nutzerbedürfnisse zugeschnitten sind, da Fachleute in verschiedenen Domänen unterschiedliche Arten von Informationen benötigen.

Feedbackloops

Ein zentrales Element humanzentrierter KI ist das Konzept des Feedbackloops. Durch Feedbackloops wird ein kontinuierlicher Austausch zwischen Mensch und Maschine geschaffen, der die Leistung und Anpassungsfähigkeit des Systems verbessert. Solche Schleifen ermöglichen es, dass die KI von menschlichen Rückmeldungen lernt und so im Laufe der Zeit optimiert wird. Ein Beispiel für die praktische Umsetzung eines Feedbackloops ist der Einsatz in der Kundendienstbranche: Wenn Kunden mit einem KI-gestützten Chatbot interagieren, können sie am Ende der Interaktion Feedback geben, ob ihre Anliegen zufriedenstellend gelöst wurden. Dieses Feedback fließt dann in die Weiterentwicklung der KI ein, sodass zukünftige Interaktionen effektiver und zielführender gestaltet werden können. Darüber hinaus können Feedbackloops helfen, Biases und Verzerrungen im System zu erkennen und zu korrigieren, was eine faire und diskriminierungsfreie Nutzung der Technologie fördert.

Welche Systeme wirken in der hKI zusammen?

In der hKI trifft die Expertise verschiedener Fachrichtungen zusammen. Die soziale und technische Dimension der Umsetzung wird begrenzt durch wirtschaftliche Voraussetzungen, die den Rahmen des in der Praxis machbaren stecken. Diese Systeme treffen sich unter anderem an den folgenden drei Stellen: in der zugrundeliegenden Technik,

Die zugrundeliegende Technik

Die technische Infrastruktur und die zugrunde liegenden Algorithmen sind zentrale Bausteine einer humanzentrierten KI. Eine solche Infrastruktur muss nicht nur sicher und zuverlässig sein, sondern auch so gestaltet werden, dass sie den spezifischen Anforderungen an Transparenz, Ethik und Datenschutz gerecht wird. Viele KI-Modelle basieren auf maschinellem Lernen und verwenden große Datenmengen zur Modellierung und Verbesserung ihrer Leistung. Eine zentrale Herausforderung ist es, sicherzustellen, dass diese Modelle so entwickelt und eingesetzt werden, dass sie den rechtlichen und ethischen Anforderungen entsprechen. Dabei sollten die Modelle nicht nur genau, sondern auch nachvollziehbar sein, sodass die Ergebnisse von Menschen überprüft und interpretiert werden können.

Interaktionspunkte von Mensch und KI

Interaktionspunkte sind die Berührungspunkte zwischen Mensch und KI, die die Kommunikation und Zusammenarbeit ermöglichen. Diese Punkte müssen so gestaltet sein, dass sie eine positive Nutzererfahrung schaffen und gleichzeitig die Kontrolle und das Vertrauen in die Technologie stärken. Beispiele für solche Interaktionspunkte sind Sprachassistenten, die auf Fragen der Nutzerinnen eingehen können, oder Diagnoseprogramme, die Ärztinnen dabei unterstützen, Gesundheitsprobleme ihrer Patient*innen schneller zu erkennen. Es ist entscheidend, dass eine ausreichende Menge der Interaktionspunkte so gestaltet sind, dass sie für den Menschen nachvollziehbar und verständlich bleiben und den Nutzern ermöglichen, die KI zu hinterfragen oder sich ihrer Entscheidungen zu widersetzen. So werden die Interaktionspunkte zu Schnittstellen.

Anforderungen an Beschäftigte

Die Einführung humanzentrierter KI in verschiedenen Branchen bringt neue Anforderungen an die Kompetenzen der Beschäftigten mit sich. Mitarbeiterinnen müssen nicht nur in der Lage sein, die KI-Systeme effektiv zu bedienen, sondern auch ein grundlegendes Verständnis für deren Funktionsweise haben. Schulungen und Weiterbildungen spielen eine wichtige Rolle, damit die Mitarbeiterinnen die Technologien sinnvoll und verantwortungsbewusst einsetzen können. Darüber hinaus sind ethische Schulungen wichtig, um die möglichen Implikationen und Herausforderungen im Umgang mit KI zu verstehen. Da die KI zunehmend in Entscheidungsprozesse integriert wird, müssen Beschäftigte in der Lage sein, die Ergebnisse kritisch zu hinterfragen und bei Bedarf zu korrigieren. Mitarbeitende sollten außerdem darin geschult sein, potenzielle Vorurteile und Fehler im System zu erkennen und entsprechende Maßnahmen zu ergreifen.

Best-Practice-Beispiele

Best-Practice-Beispiele geben einen praxisnahen Einblick, wie humanzentrierte KI in verschiedenen Branchen sinnvoll eingesetzt werden kann, und was aus der Perspektive der hKI in diesen Szenarien zu beachten ist.

Medizinische Diagnostik:
In der medizinischen Diagnostik beispielsweise unterstützen KI-Systeme Ärztinnen bei der Auswertung von Bilddaten wie Röntgenaufnahmen oder MRT-Scans. Diese Systeme können Anomalien in den Bildern erkennen und so Ärztinnen Hinweise auf mögliche gesundheitliche Probleme wie Frühsignale einer Alzheimer-Erkrankung geben, was die Diagnose beschleunigt und präzisiert.
Andere Beispiele, wie personalisierte Begleitung von Krebspatient:innen mithilfe eines digitalen Symptom-Tagebuchs oder KI-basiert regulierte Insulin-Pens sollen aus hKI-Perspektive in enger Zusammenarbeit mit den medizinischen Fachkräften umgesetzt werden. Dabei bleibt die Entscheidung über die Diagnose und Behandlung jedoch stets in der Hand der Fachleute. Diese Art von KI-Einsatz zeigt, wie die Technologie als Unterstützung und nicht als Ersatz dient und das Wohl der Patienten in den Vordergrund stellt.

Bildung:
Ein weiteres Beispiel ist die Anwendung von KI in der Bildung. Adaptive Lernsysteme analysieren den Fortschritt von Schüler*innen und passen die Lerninhalte individuell an deren Bedürfnisse an. Ein solches System kann erkennen, welche Themen die Lernenden bereits verstanden haben und bei welchen sie noch Unterstützung benötigen. Hierdurch wird der Lernprozess effizienter und zielgerichteter, was besonders in großen Schulklassen mit heterogenen Leistungsniveaus von Vorteil ist. Diese Systeme tragen dazu bei, die Chancengleichheit zu fördern, da sie eine gezielte Förderung ermöglichen.
So können Chatbots in Echtzeit Vorschläge für die nächsten Lernschritte machen, Illustrationen für schwer verständliche Themen erstellen, alternative Lösungswege aufzeigen oder Querverbindungen mit anderen Lehrinhalten verdeutlichen. Lehrer:innen und Schüler:innen sind in diesem Szenario zu betrachten. Sind die Lehrer realistisch in der Lage, die KI-Vorschläge für all ihre Schüler:innen zu überwachen und Änderungen anhand ihrer Einschätzung vorzunehmen? Vertrauen die Schüler:innen den Vorschlägen der KI? Das Vertrauen in die Technologie ist bei allen Beteiligten aufzubauen. Selbst im schulischen Kontext ist eine Orientierung an den Schritten des Change Managements zu empfehlen.
Mehr zu diesem Szenario ist in einem Diskussionspapier von KI-Campus zu erfahren.

Industrie 4.0 und Automatisierung:
In der Industrie und Fertigung wird KI genutzt, um Prozesse effizienter und sicherer zu gestalten. Ein Beispiel ist der Einsatz von KI zur Überwachung und Steuerung von Produktionsprozessen. Diese Systeme können Abweichungen erkennen, bevor sie zu größeren Problemen führen, und so die Produktion optimieren. Durch den Einsatz von KI können Beschäftigte von Routineaufgaben entlastet werden und sich auf kreative und strategische Aufgaben konzentrieren. Auch hier zeigt sich der Nutzen humanzentrierter KI, indem sie den Arbeitsalltag der Menschen erleichtert und zur Wertschöpfung beiträgt.

Literatur

Bratan, Tanja (2023): Künstliche Intelligenz im Gesundheitsbereich: Eine Übersicht. https://www.isi.fraunhofer.de/de/blog/2023/kuenstliche-intelligenz-im-gesundheitsbereich.html (Abruf am 27.11.2024).

Harbarth, Lydia/Bodemer, Daniel/Schnaubert, Lenka (2023). Elaborateness of explanations to understand AI recommendations. In Blikstein, P., Van Aalst, J., Kizito, R., & Brennan, K. (Eds.), *Proceedings of the 17th International Conference of the Learning Sciences - ICLS 2023 (pp. 1827-1828)*. International Society of the Learning Sciences. https://doi.org/10.22318/icls2023.373839.

Huchler, Norbert/ Adolph, Lars/ André, Elisabeth; Bauer et. al. (2020): Kriterien für die Mensch-Maschine- Interaktion bei KI. Ansätze für die menschengerechte Gestaltung in der Arbeitswelt. Hg. v. Plattform Lernende Systeme. https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG2_Whitepaper2_220620.pdf (Abruf am 28.11.2024).

Lepri, Bruno/Nuria, oliver/Pentland, Alex (2021): Ethical machines: The human-centric use of artificial intelligence. In: iScience 24 (3). https://www.sciencedirect.com/science/article/pii/S2589004221002170 (abruf am 12.11.2024).

Schleiss, Johannes/ Mah, Dana-Kristin/ Böhme, Katrin et. al. (2023): Künstliche Intelligenz in der Bildung. Drei Zukunftsszenarien und fünf Handlungsfelder. https://ki-campus.org/sites/default/files/2023-03/2023-03_Diskussionspapier_KI_Bildung_Zukunftsszenarien_Handlungsfelder_KI-Campus_1.pdf (Abruf am 28.11.2024).

Sheikh, H./ Prins, C./ Schrijvers/ E. (2023): Artificial Intelligence: Definition and Background. In: Mission AI. Research for Policy. Springer, Cham. https://doi.org/10.1007/978-3-031-21448-6_2 (Abruf am 12.11.2024).