Zusammenspiel von CDR und EU AI Act bei menschenzentrierter KI

Oliver Merx

1. Menschenzentrierung im digitalen Wandel

1.1 Begriffsevolution seit den 80er Jahren

Der Bundesverband der Deutschen Industrie (BDI) schreibt auf seiner Website:

> Menschenzentriertes Handeln ist ein neues Paradigma, das beim technologischen Wandel mitgedacht werden muss [...]. Die Wirtschaft wird nicht mehr einer rein quantitativen Logik folgen [...]. Die „Customer Centricity" weicht der „Human Centricity" – und Produkte, Prozesse und Unternehmensentscheidungen spiegeln diese wider (BDI 2024).

Es geht um das Ziel, Menschen umfassend in den Mittelpunkt zu stellen: Bei Produkten und Services als auch bei Organisationen und Prozessen. Der BDI bringt das Thema auf den Punkt. Mit einer Ausnahme: Menschenzentrierung ist im Kontext von IT und Technik nicht wirklich neu. Die Wurzeln der „Human Centricity" reichen bis in die 80er Jahre zurück. Damals wurde u. a. das „User Centered Design" (UCD) entwickelt. Dessen Ziel lag und liegt darin, die Mensch-Maschine-Interaktion zu verbessern – allerdings vornehmlich, um ökonomische Vorteile zu erzielen. Daraus entwickelte sich im Lauf der Zeit das „Human Centered Design" (HCD), welches in der ISO 9241-210: 2019 (vgl. DIN 2019) manifestiert wurde. Adaptiert wurde das Prinzip der Menschenzentrierung sukzessive auch im Business (vgl. Accenture 2020). So betrachtet ist vergleichsweise neu, dass sich die „Human Centricity" nicht mehr primär auf haptische Produkte und Software, sondern auch auf Digitale Services, Prozesse und Organisationen bezieht.

Bedeutender ist jedoch die Tatsache, dass sich selbst in der Wirtschaft über die ökonomischen Vorteile hinaus ein stark ethisch geprägtes Verständnis der „Human Centricty" herausgemendelt hat. Diese Entwicklung wurde maßgeblich von der digitalen Transformation, ihrem exponentiellen Innovationstempo und mitwachsenden Risiken beeinflusst. Eine Entwicklung, die zu steigender Angst vieler Menschen vor Missbrauch, Kontroll- und Arbeitsplatzverlust führte.

Vor dem Hintergrund der skizzierten Entwicklung umfasst digitale Menschenzentrierung heute einen Mix aus eher ökonomisch und überwiegend ethischen geprägten Aspekten. Häufig genannt werden u. a. folgende Aspekte:

- Einbeziehung von Menschen in Gestaltungsprozesse,
- UX und gute Bedienbarkeit,
- Prüfbarkeit und Kontrolle durch Menschen,
- Förderung digitaler Kompetenz,
- Schutz vor Diskriminierung und Manipulation,
- Souveränität u. a. bzgl. persönlicher Daten und Entscheidungen,
- Sicherheit und Schutz der Gesundheit,
- Transparenz von technischen Abläufen und Entscheidungsprozessen,
- Nachhaltigkeit und Schutz der Umwelt,
- Individuelle Beschwerderechte (auch gegenüber dem Staat).

Die vorherigen Aspekte lassen sich auf einer „Landkarte" mit direkter und indirekter sowie ökonomischer und/oder ethischer Menschenzentrierung strukturieren. Dabei sind die meisten Kriterien nicht primär auf KI bezogen: Sie gelten im gesamten Bereich von Digitalisierung und Automatisierung, also auch für „KI-freie" Webseiten, Online-Shops, (Mobile-)Apps sowie sonstige Services und digital gesteuerte Produkte wie Roboter und Automaten.

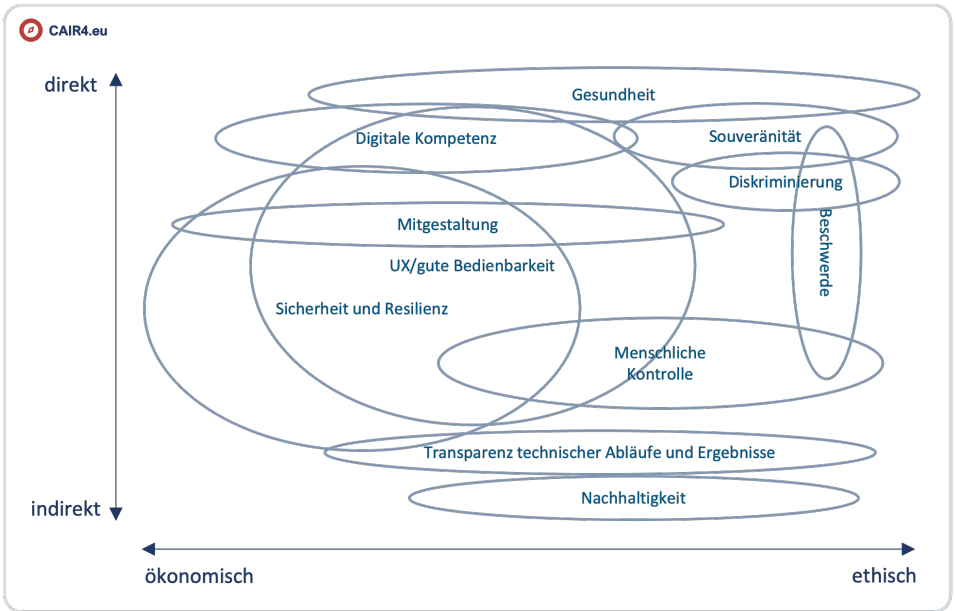

ABBILDUNG 1: LANDKARTE ZUR DIGITALEN MENSCHENZENTRIERUNG (QUELLE: EIGENE ABBILDUNG)

Nahezu alle der skizzierten Aspekte finden sich im EU AI Act wieder. Doch bevor dieser realisiert werden konnte, waren sie zunächst vor allem freiwillige Handlungsfelder der Corporate Digital Responsibility (CDR). Umso aufschlussreicher ist die Unterteilung der einzelnen Aspekte im Hinblick auf überwiegend ethische bzw. ökonomische Schwerpunkte:

- Die Motivation, digitale Menschenzentrierung freiwillig und eigeninitiativ zu beachten und umzusetzen, hängt in einer zahlengeprägten Wirtschaft wesentlich davon ab, ob damit auch ökonomische Vorteile erzielt werden können.
- Man fand und findet man daher im Kontext der CDR häufig eine Kombination von Maßnahmen, die das „ethisch Angenehme“ mit dem „ökonomisch Nützlichen“ im Sinne eines Kombipakets verbinden. Idealerweise kommt noch mediale Awareness hinzu, frei nach dem Motto „Tue Gutes und rede darüber!“ Das ist auch gut so.
- Gleichwohl macht es einen Unterschied, ob eine Organisation im Sinne der CDR ausschließlich dort freiwillig aktiv wird, wo es geringe Überwindung und wenig Aktivität, dafür aber

einfach zu erzielenden Imagegewinn plus Wettbewerbsvorteile verspricht bzw. das Engagement gezielt dort unterlässt, wo diese Kombination nicht gegeben ist, also keine „Opfer“ erbracht werden müssen.

- Noch kritischer ist es, wenn nur so getan wird „als ob“. Insofern sind „ethische Fassaden“ und „Greenwashing“ (vgl. Mütze 2022) die oft kritisierte Kehrseite aller auf Freiwilligkeit beruhenden Ansätze.

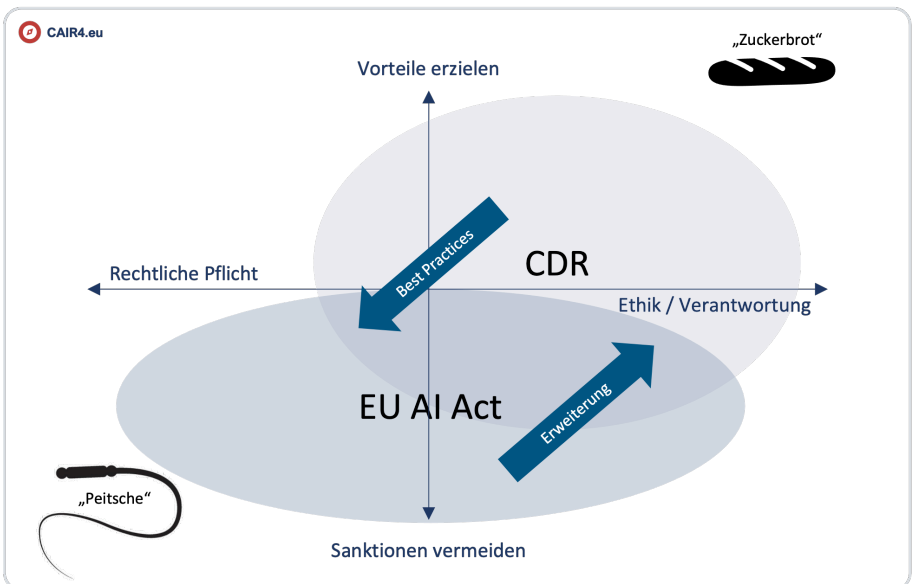

ABBILDUNG 2: SYMBIOSE VON CDR UND EU AI ACT
(QUELLE: EIGENE ABBILDUNG)

Vor diesem Hintergrund lässt sich das Zusammenspiel von freiwilliger Verantwortungsübernahme im Sinne der CDR und gesetzlicher Pflicht durchaus mit dem Gleichnis von „Zuckerbrot und Peitsche“ skizzieren – dabei ist ersteres noch attraktiver, wenn letztere droht! Das Verhältnis kann im Sinne eines Stufenmodells konstruktiv-symbiotisch beschrieben werden:

1. Freiwilliges Engagement und CDR sind der schnelllebigen Welt der Digitalisierung der ideale „First Mover". Sie bahnen den Weg, um wertvolle Erfahrungen zu sammeln. Wer früh agiert, wird nicht nur mit Imagegewinn belohnt: Wer Erfahrungen aufbaut und teilt wird zum aktiven Mitgestalter.
2. Sobald die Freiwilligkeit in Fläche auf der Stelle tritt, muss der nächste Schritt darin liegen, verbindliche gesetzliche Vorgaben zu evaluieren. Insbesondere bei hoch komplexer KI ist dies ohne die umfassende Berücksichtigung anerkannter Best Practices der CDR kaum vorstellbar. Insofern zahlt sich in der zweiten Stufe das freiwillige Engagement der ersten Stufe noch einmal aus.
3. Demjenigen, der in den ersten beiden Phasen nur zögerlich oder gar nicht agiert, dem drohen in Phase drei signifikante Nachteile: Mitunter müssen Investitionen abgeschrieben und Prozesse komplett neugestaltet sowie hoch spezialisierte Mitarbeiter und Partner gewonnen werden.

Damit die wichtigen „First-Mover-Advantages" in der ersten und zweiten Stufe plausibel erzielt werden können, scheint es geradezu denknotwendig, dass es irgendwann zur dritten Phase rechtlicher Regulierung kommt. Drohende Sanktionen inbegriffen! Im Hinblick auf die Menschenzentrierung von KI ist dieser Fall nun eingetreten. Zu Recht!

1.3 Ohne Druck geht es nicht

Ein „motivationsfördernder Druck", um im Daten-, KI-Umfeld „freiwillig" mehr zu tun als explizit vorgeschrieben, entstand zunächst durch das zunehmende Risiko negativer Medienberichte (vgl. Phoenix 2018). Doch selbst in solchen Fällen wurde und wird häufig ausschließlich reaktiv gehandelt – mitunter so lange, bis sich der Missbrauch nicht mehr leugnen lässt! Der gesellschaftliche Vertrauensschaden wächst dabei mit jedem aufgedeckten, erst recht mit jedem zu Unrecht bestrittenen Missbrauch weiter an. Im Nachgang ist das erodierte kollektive Vertrauen kaum mehr wieder herzustellen. Folglich muss es beim EU AI Act primär darum gehen, die Verwirklichung bekannter KI-Risiken so weit wie möglich präventiv zu verhindern – auch im Wege der „Abschreckung".

Eine vergleichsweise softe Variante stellen im Vergleich dazu die Berichtspflichten der Corporate Social Responsibility (CSR)[1] dar. Sie haben den Druck des Nachweises (präventiver) Aktivität wahrnehmbar erhöht. Zudem fließen immer öfter digitale Themen in CSR-Berichte ein – nicht zuletzt beim Energieverbrauch und sozialen Aspekten wie digitaler Inklusion.

Der Trend, digitale Themen wie vertrauenswürdige Datennutzung, ethische KI und Energieeffizienz ins CSR-Reporting zu integrieren, hat jedoch der Breite der Gesellschaft kaum zur Reduktion von Skandalen oder mehr Vertrauen in KI geführt. Mancher Insider argwöhnt sogar, dass es bei entsprechenden Berichten häufig nur darum gehe, Investoren zu ködern (vgl. Breinich-Schilly 2023).

1.4 CDR als „First Mover" menschenzentrierter KI

Insofern war und ist es von Bedeutung, dass digitale Menschenzentrierung im ethischen Sinne von Beginn an eine der tragenden Säulen sowohl des CDR-Kodex des BMUV als auch der CDR-Building-Bloxx (vgl. Whyzer 2021)gewesen ist: Beide Frameworks beruhen auf Freiwilligkeit. Bei beiden Initiativen darf und soll glaubhaftes ethisches Engagement von ökonomischem Nutzen sein und Wettbewerbsvorteile ermöglichen. Zwar führen beide Initiativen (ebenso wie die DRGs (vgl. Identity Valley 2024) oder das internationale CDR-Manifesto (vgl. CDR 2024)) trotz jahrelangem Engagements eher ein Nischendasein – umso mehr können sie als Brückenbauer im Rahmen des Stufenmodells betrachtet werden!

CDR ist (wenngleich von vielen unbemerkt) ein wichtiger Geburtshelfer des EU AI Acts gewesen: Das gilt für die Evaluation von Best Practices ebenso wie für das Ausarbeiten von Verhaltenskodizes oder das Erstellen fachlicher Leitlinien. Ohne die Vielzahl entsprechender Erfahrungswerte der digitalen Praxis hätte der EU AI Act mit hoher Wahrscheinlichkeit erst viel später das Licht der Welt erblicken können. In diesem Sinne beruht er in vieler Hinsicht auf dem Input engagierter Unternehmen, NGOs sowie von Ethik- und Digital-Experten, eben auf CDR!

Damit zum Status quo: Mit dem Inkrafttreten des EU AI Acts am 2. August 2024 wird die Menschenzentrierung von KI zum rechtlich verpflichtenden Inhalt erhoben. Sie stellt sogar das zentrale Schutzgut dar, dass für alle Akteure im Rahmen der KI-Lieferkette verbindlich ist.

---

[1] Siehe u. a.: https://www.csr-in-deutschland.de und https://sdg-portal.de.

2. Verankerung digitaler Menschenzentrierung im EU AI Act

2.1 Initiale Hervorhebung in der Begründung des EU AI Acts

Bereits die erste Ziffer der Begründung des EU AI Acts stellt klar:

> Zweck dieser Verordnung ist es [...] die Einführung von menschenzentrierter und vertrauenswürdiger künstlicher Intelligenz (KI) zu fördern und gleichzeitig ein hohes Schutzniveau in Bezug auf Gesundheit, Sicherheit und der in der Charta der Grundrechte der Europäischen Union („Charta") verankerten Grundrechte, einschließlich Demokratie, Rechtsstaatlichkeit und Umweltschutz, sicherzustellen (Europäisches Parlament / Rat der Europäischen Union 2024).

Prominenter als im ersten Satz der gut 800 Seiten umfassenden Norm kann das Ziel der Menschenzentrierung vermutlich kaum integriert werden. Ähnlich wie bei der KI-Strategie des Bundes (vgl. Bundesregierung 2018) steht die vertrauensfördernde Menschenzentrierung bewusst neben anderen, explizit genannten Schutzgütern wie Gesundheit, Sicherheit und Umweltschutz.

Diese Klarstellung ist wichtig, da die anderen Aspekte nur teilweise auf dem Ziel direkter Menschenzentrierung beruhen. Gleichwohl gibt es fließende Übergänge:

- So ist der Schutz der Gesundheit ohne Frage ein Aspekt, welcher einer direkten ethischen Zielsetzung im Hinblick auf Menschenzentrierung entspringt.
- Beim Thema Sicherheit ist dies nur teilweise der Fall, da hier neben der Sicherheit für Menschen auch der Schutz und die Resilienz kritischer Infrastrukturen gemeint ist.
- Ähnlich ist es beim Umweltschutz, der ebenso wie Nachhaltigkeit nicht primär Ausdruck menschenzentrierter Vertrauensförderung, sondern ein eigener unabhängiger Wert ist.

Diese Betrachtungsweise wird durch zwei weitere Ziffern der Begründung unterstützt. So heißt es in Ziffer (6) der KI-Verordnung:

> Angesichts der großen Auswirkungen, die KI auf die Gesellschaft haben kann, und der Notwendigkeit, Vertrauen aufzubauen, ist es von entscheidender Bedeutung, dass KI [...] eine menschenzentrierte Technologie ist. Sie sollte den

> Menschen als Instrument dienen und letztendlich das menschliche Wohlergehen verbessern (Europäisches Parlament / Rat der Europäischen Union 2024).

Ziffer (8) ergänzt die Eigenständigkeit und den Vertrauensfokus:

> Durch die Festlegung dieser Vorschriften [...] unterstützt diese Verordnung das vom Europäischen Rat formulierte Ziel, das europäische menschenzentrierte KI-Konzept zu fördern und bei der Entwicklung einer sicheren, vertrauenswürdigen und ethisch vertretbaren KI weltweit eine Führungsrolle einzunehmen [...] (ebd.).

2.2 Verankerung in Artikel 1 EU AI Act

Wichtiger noch als in der Begründung ist die unmittelbare Verankerung der Menschenzentrierung in Artikel 1 EU AI Act. Dort heißt es: „Zweck dieser Verordnung ist es, das Funktionieren des Binnenmarkts zu verbessern und die Einführung einer auf den Menschen ausgerichteten und vertrauenswürdigen künstlichen Intelligenz (KI) zu fördern." Die Formulierung „auf den Menschen ausgerichtet" entspricht inhaltlich voll und ganz der in der Begründung verwendeten Formulierung „Menschenzentrierung". Besonders deutlich wird dies in der englischen Version, die in der Begründung wie dem Normtext die Formulierung „human-centric" verwendet.

Nichtsdestotrotz bleibt eine elementare Frage offen: Was genau ist mit „Menschenzentrierung" bzw. der „Ausrichtung auf den Menschen" konkret gemeint? Die Frage kommt nicht von ungefähr, denn Tatsache ist, dass trotz des Umfangs der KI-Verordnung von rund 800 Seiten:

- Eine eindeutige Definition, wie sie in Artikel 3 EU AI Act für immerhin 68 andere wichtige Begriffe im KI-Kontext existiert, für die „Human Centricity" fehlt.
- Die Menschenzentrierung des Artikel 1 EU AI Act wird damit in vieler Hinsicht zum unbestimmten und auslegungsbedürftigen Rechtsbegriff.
- Dies macht eine zeitbezogene Interpretation der Menschenzentrierung im KI-Kontext einerseits möglich, andererseits macht es diese auch stets aufs Neue erforderlich.

An dieser Stelle gewinnt die anfangs skizzierte Begriffsevolution sowie die mittels CDR, der Wissenschaft und anderer anerkannter Instrumente geschaffene Bandbreite der unterschiedlichen Inhalte von Menschenzentrierung an Bedeutung. Sie helfen dabei, die einzelnen Normen des EU AI Acts dahingehend zu scannen, welche Aspekte der „Human Centricity" wie gesetzlich geregelt werden.

3. Normen des EU AI Acts, die Menschenzentrierung konkretisieren

3.1 Gestaffelte Menschenzentrierung im EU AI Act

Die in Abbildung 1 initial skizzierte Landkarte digitaler Menschenzentrierung dient dazu, um im Rahmen des EU AI Acts in gestaffelter Form nach solchen Normen zu suchen, welche sich auf „Human Centricity" beziehen. Dabei sind nach der hier vertretenen Auffassung drei Varianten zu unterscheiden:

- Normen mit begleitender Konkretisierung (gepunktete Linie),
- Artikel, die Menschenzentrierung verordnungsübergreifend regeln (gestrichelte Linie)
- und spezifische Normen, die besonders wichtige Punkte regeln (durchgezogene Linie)

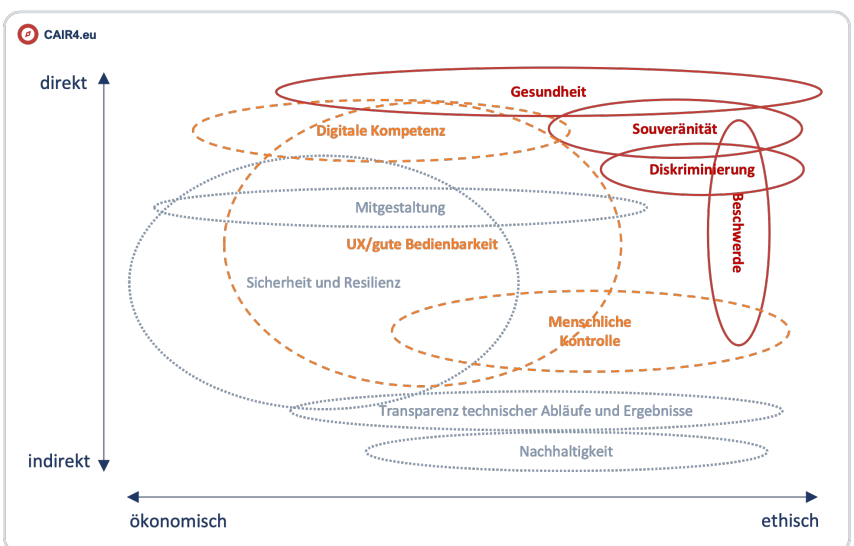

ABBILDUNG 3: GESTAFFELTE MENSCHENZENTRIERUNG IM EU AI ACT (QUELLE: EIGENE ABBILDUNG)

Sicher gibt es noch weit mehr Inhalte und Möglichkeiten, die Aspekte zu strukturieren, doch schon dieses vergleichsweise grobe Modell ermöglicht aufschlussreiche Einblicke.

3.2 Normen mit begleitender Konkretisierung von Menschenzentrierung

„Nachhaltigkeit", „Transparenz", „Sicherheit" sowie „Mitgestaltung" sind Aspekte, die nach der hier vertretenen Auffassung die Menschenzentrierung von KI eher begleitend konkretisieren, da bei ihnen der Übergang zu anderen selbständigen Schutzgütern fließend erscheint bzw. deren Fokus zum Teil auch überwiegt.

- Nachhaltigkeit und Energieeffizienz werden u. a. in den Artikeln 95 (2) b), 112 (7) sowie Anhang XIII c) EU AI Act als auch den Ziffern (4) und (174) der Begründung behandelt. Energieeffizienz kommt aber in Anbetracht steigender Energiekosten auch der Wirtschaft sowie der Gemeinschaft als Ganzes zugute. Beides sind Werte mit eigenem Schutzbereich, haben aber durchaus Bezug zur Menschenzentrierung.
- Die IT- bzw. Cyber-Sicherheit findet sich in nicht weniger als elf Ziffern der Begründung (54, 55, 74, 76, 77, 78, 114, 115, 122, 126, 131), neun Artikeln (13, 15, 31, 42, 55, 59, 66, 70, 78) sowie in Anhang IV wieder. Im Detail wird sie u. a. im Hinblick auf Hochrisiko-KI und KI-Modelle thematisiert (auch in Kombination mit dem Cyber Resilience Act (CRA) (Europäische Kommission 2022). Entsprechende Informationen sind u. a. aufgrund der fachlichen Detailtiefe wohl eher ein indirektes Instrument zur kollektiven Vertrauensförderung bzw. Menschenzentrierung.
- Das Transparenzerfordernis zieht sich durch mehrere Kapitel des EU AI Acts: Erforderlich ist es sowohl für KI-Systeme hoher und mittlerer Risiken als auch für KI-Modelle mit allgemeinem Verwendungszweck. Für beide ist die Transparenzpflicht z.T. sehr detailliert geregelt. Als Normen zu nennen sind u. a. Artikel 13, 50, 51 ff. und 96 sowie Anhang XII EU AI Act. Hinzu kommen über ein Dutzend Vertiefungen in der Begründung, insbesondere in Ziffer (107), (135) und (137). Obwohl Transparenz häufig als eines der wichtigsten Merkmale vertrauenswürdiger KI betrachtet wird, ist sie im EU AI Act nur an vergleichsweise wenigen Stellen für Endnutzer bedeutsam, z. B. bei KI-Chatbots, welche gemäß Artikel 50 EU AI Act als solche erkennbar sein müssen. Viele andere Aspekte der Transparenz beziehen sich hingegen auf Pflichten in der KI-Lieferkette, weshalb sie hier als begleitende Konkretisierung der Menschenzentrierung eingestuft wird.

- Auch zur Mitgestaltung finden sich Inhalte im EU AI Act, so z. B. Ziffer (20) im Hinblick auf die Validierung von Konzepten oder Ziffer (165) zur Beteiligung verschiedener Interessensträger bis hin zur ausgewogenen Beteiligung unterschiedlicher Geschlechter. Dieser Punkt besitzt ebenfalls starke Überschneidungen mit der Menschenzentrierung von KI. Diesbezügliche Normen sind aber nach der hier vertretenen Auffassung vor allem für KI-Experten von verfahrensbezogener Relevanz. Kollektive Vertrauensbildung erscheint dabei – je nach Perspektive – eher als Nebeneffekt. Insofern erfolgt auch hier eine vergleichsweise begleitende Konkretisierung.

In Anbetracht der Fülle von Inhalten des EU AI Acts, die alle auf irgendeine Weise mit Menschenzentrierung in Zusammenhang stehen, wird bereits an dieser Stelle deutlich, dass der EU AI Act viele Regelungen enthält, um den unbestimmten Rechtsbegriff der „Human Centricity" im jeweiligen Kontext zumindest begleitend zu konkretisieren.

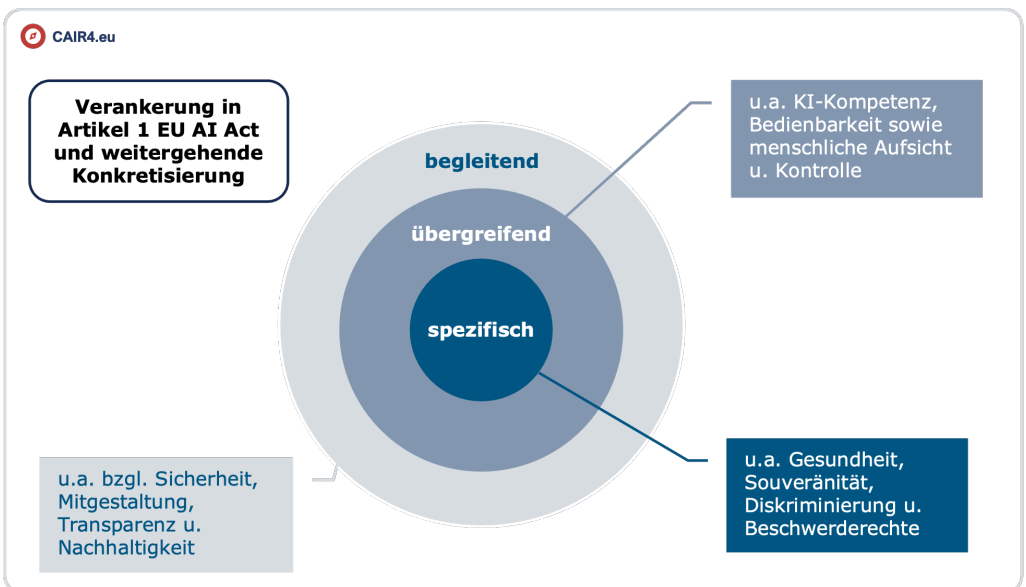

ABBILDUNG 4: GESTAFFELTE MENSCHENZENTRIERUNG IM EU AI ACT (QUELLE: EIGENE ABBILDUNG)

Vor diesem Hintergrund werden nachfolgend jene Aspekte etwas genauer betrachtet, welche die Menschenzentrierung normübergreifend fokussierter konkretisieren. Darunter „KI-Kompetenz", „Bedienbarkeit" und „menschliche Kontrolle".

KI-Kompetenz

Sie wird in Artikel 4 EU AI Act als Teil des ersten Kapitels „vor die Klammer gezogen": Demnach müssen sowohl die Anbieter als auch die Betreiber von KI-Systemen Maßnahmen ergreifen, um sicherzustellen, dass Personen, die mit dem Betrieb oder der Nutzung von KI-Systemen befasst sind, über ein ausreichendes Maß an KI-Kompetenz verfügen. Wichtig ist, dass hier nicht zwischen unterschiedlichen Risikoarten unterschieden wird: Es ist zwar keine sanktionierbare Pflicht, jedoch eine justiziable Leitlinie, die selbst bei Systemen mit geringen Risiken zu erfüllen ist (vgl. Merx 2024).

Aufschlussreich ist diesbezüglich u. a. die in Ziffer (165) enthaltene Anregung, dass bei KI mit geringen Risiken Verhaltenskodizes erstellt werden sollten, in denen dargelegt wird, wie man das Kriterium der KI-Kompetenz umzusetzen gedenkt. Beispielsweise sollten besonders schutzbedürftige Personen im Hinblick auf Barrierefreiheit gefördert werden. Wie man sieht: Der EU AI Act thematisiert und fördert an vielen Stellen CDR – dazu mehr am Ende.

Noch weiter geht Ziffer (20). Darin wird deutlich, welch hohe Bedeutung der KI-Kompetenz im Rahmen von KI-Wertschöpfungsketten zukommt. Es geht insbesondere darum, dass alle Akteure über den „eigenen Tellerrand" hinauszublicken. Vertrauensbildende Menschenzentrierung hört eben nicht vor der eigenen Haustür auf! Vielmehr erfordert sie ein hohes Maß an Kooperation von allen an der Entwicklung, Inverkehrbringung und Nutzung von KI beteiligten Akteuren. Interdisziplinarität ist wiederum ein wichtiger Aspekt der CDR – diese ergänzt auch in diesem Fall die gesetzlichen Pflichten, da der EU AI Act bewusst offenlässt, wie genau die Umsetzung im Einzelfall auszugestalten ist.

Bedienbarkeit

Der zuvor skizzierte Punkt genereller KI-Kompetenz wird an verschiedenen Stellen noch einmal deutlich tiefer gelegt: Während die allgemeine Regelung des Artikel 4 EU AI Act ein umfassendes Fundament bildet, formulieren Regelungen wie Artikel 15, 23, 24, 26, 27, 60 sowie Anhang IV, VIII, XI und XII EU

AI Act durchaus detaillierte Vorgaben für die Notwendigkeit und Ausgestaltung von Betriebsanleitungen, um die bestmögliche Befähigung zur Bedienung (insbesondere von Hochrisiko-KI-Systemen) zu gewährleisten.

Es reicht also nicht aus, zur Kontrolle der Risiken von KI eine vermeintlich intuitiv bedienbare Oberfläche zu gestalten und die Nutzer sich selbst zu überlassen. Vielmehr werden Anbieter von KI-Systemen ebenso wie von KI-Modellen verpflichtet, ausführliche Bedienungskonzepte zu liefern sowie die erforderlichen Trainings zu ermöglichen – und nachzuweisen!

Bei Hochrisiko-KI i.S.v. Artikel 6 ff. EU AI Act gehört dazu auch die sich aus Artikel 26 EU AI Act ergebende Pflicht der Betreiber bzw. Nutzer, wichtige Information zu den Anbietern zurückzuspielen, also einen Informationskreislauf zu bilden.

Werden die zuvor genannten Pflichten nicht erfüllt, drohen den verschiedenen Akteuren Sanktionen. Gemäß Artikel 99 EU AI Act können diese durchaus empfindlich ausfallen – der Zweck liegt gemäß Absatz 1 u. a. darin, abzuschrecken! Entsprechend Absatz 4 e) und g) drohen im Einzelfall sogar Nutzern Sanktionen. Da die „Peitsche" des EU AI Acts präventiv wirken will und muss, erscheint dies durchaus angemessen.

Menschliche Aufsicht und Kontrolle

Die zuvor genannten Vorgaben sind Voraussetzung dafür, dass der Mensch weitgehend die Kontrolle über KI erhält bzw. behält. Im Hinblick auf Hochrisiko-KI geht Artikel 14 EU AI Act noch einen Schritt weiter, in dem er Details zur menschlichen Aufsicht regelt – was u. a. dazu führt, dass die Akteure im Wege eines „red buttons" theoretisch in der Lage sein müssten, eine „aus dem Ruder laufende KI" während des Betriebs zu stoppen (Fyler 2023). In eine ähnliche Richtung geht Artikel 72 EU AI Act, der von Anbieter einer Hochrisiko-KI einen Plan zur Beobachtung über den gesamten Life Cycle des Systems einfordert. Ein wichtiger Punkt, um allen Akteuren das Vertrauen zu vermitteln, dass menschenzentrierte KI kein Sprint, sondern ein Marathon-Lauf ist.

Ob und wie weit die Vorgaben der Artikel 14 und 72 EU AI Act tatsächlich in der Praxis so wie geregelt umgesetzt werden können, erscheint aktuell noch offen. Insbesondere bei selbstlernender KI ist es mitunter sehr schwer, den Lernerfolg bzw. die diesbezüglichen Fortschritte eines Systems effektiv zu beaufsichtigen oder zu dokumentieren. Schließlich sind viele der besonders leistungsfähigen KI-Systeme und KI-Modelle (z. B. im Fall von Deep Learning) eine „Black Box".

Zugleich ist der Weg zu einer „explainable AI“ (XAI), welche umfassende Kontrolle zu ermöglichen verspricht, nach aktuellen Experteneinschätzungen noch lang (vgl. Eviden 2024).

Dies wiederum wirft die Frage auf, warum der EU AI Act mitunter recht detaillierte Vorschriften enthält, die fachlich als auch ökonomisch kaum umsetzbar erscheinen. Das nicht zuletzt für das Ziel größtmöglicher Transparenz. Diese kann durchaus zu einem Bumerang-Effekt führen: Je größer die Transparenz einer KI, desto mehr steigt auch die Gefahr, dass diese manipuliert werden kann (vgl. Carli et al. 2022). Insofern ist nicht nur fraglich, ob und wie weit dieses Ziel in der Praxis erreichbar ist, sondern auch, in welchem Umfang es wirklich sinnvoll erscheint.

So oder so: Die drei zuvor skizzierten übergreifenden Aspekte „KI-Kompetenz“, „Bedienbarkeit“ und „menschliche Aufsicht“ tragen wesentlich dazu dabei das in Artikel 1 EU AI eher unbestimmt erscheinende Ziel der Menschenzentrierung in einzelnen Normen Use-Case-gerecht zu konkretisieren.

3.4 Spezifische Regelungen des EU AI Acts im Hinblick auf Menschenzentrierung

Nun zu einigen besonders spezifisch erscheinenden Aspekten, in denen Menschenzentrierung konkretisiert wird: Gemeint sind der „Schutz der Gesundheit“, die „Wahrung der Souveränität“, der „Schutz vor Diskriminierung“ sowie das individuelle „Recht auf Beschwerde“ bei mutmaßlichen Verstößen gegen den EU AI Act.

Schutz der Gesundheit / Einstufung als Hochrisiko-KI

Die Gesundheit ist ein höchstpersönliches Schutzgut. Im Sinne der Artikel 2, 3, 35 EU-Charta sowie Artikel 2 (2) Grundgesetz steht sie zudem unter besonders hohem Schutz.

Nicht ohne Grund ist daher ein großer Teil medizinischer KI, welche meist auch der Medical Device Regulation (MDR) (Europäisches Parlament / Rat der Europäischen Union 2017) unterliegt, als Hochrisiko-KI zu klassifizieren. Dies ergibt sich aus Artikel 3 (2) i. V. m. Anhang III EU AI Act (eingeschränkt durch Artikel 3 (3) EU AI Act, u. a. dann, wenn die KI eine Entscheidungsfindung nur unwesentlich beeinflusst; ebenfalls eingeschränkt durch Artikel 2 (6) EU AI Act im Falle ausschließlicher Forschung).

Zum Schutz der Gesundheit kann sogar ein bereits zertifiziertes KI-System von einer Behörde nachträglich „ausgebremst“ werden, falls von diesem i. S. v. Artikel 82 EU AI Act ein zuvor

unerkanntes Risiko ausgehen sollte. Umgekehrt können zur Wahrung der Gesundheit Zertifizierungsverfahren für KI beschleunigt werden (siehe Artikel 46 EU AI Act sowie Ziffer 130 der Begründung).

Dies ist z. B. im Fall einer neuen Epidemie vorstellbar. Menschenzentrierung im Sinne eines durch KI ermöglichten Gesundheitsschutzes bedeutet somit auch, die Flexibilität ihrer Nutzung in Notfällen zu erhöhen – ein wichtiger Aspekt, um in begründeten Einzelfällen die Chancen von KI zur Wahrung und Förderung von Gesundheit nutzen zu können.

Schutz der Souveränität

Ein souveräner Mensch ist in der Lage, sein Denken und Handeln eigenverantwortlich und selbstbestimmt zu gestalten. Bei KI ist dies häufig nur eingeschränkt möglich:

- Zum Beispiel, dann wenn man gar nicht bemerkt, dass das eigene Leben von KI mitbestimmt bzw. manipuliert wird (u. a. bei Fake Inhalten, automatisierten Bewertungen von Leistungen oder Targeting-Profilen im Marketing),
- oder falls man es bemerkt, die konkreten Auswirkungen für die eigene Person nicht erkennt bzw. die dahinter liegenden Zusammenhänge nicht genau versteht,
- oder wenn man am Ende so oder so nichts dagegen tun kann, selbst wenn man die Auswirkungen erkennt und versteht – es also an Rechtmitteln fehlt.

Um den Schutz der Souveränität im Hinblick auf KI gewährleisten zu können, reicht es vor diesem Hintergrund nicht aus, primär auf die Eigenverantwortlichkeit des einzelnen Menschen zu setzen, damit dieser sich selbst vor KI-Missbrauch schützt.

In Ziffer 133 der Begründung des EU AI Acts heißt es dazu im Hinblick auf Fake-Inhalte: „Eine Vielzahl von KI-Systemen kann große Mengen synthetischer Inhalte erzeugen, bei denen es für Menschen immer schwieriger wird, sie vom Menschen erzeugten und authentischen Inhalten zu unterscheiden“ (Europäisches Parlaments /Rat der Europäischen Union 2024). Noch schwieriger ist es bei Systemen zur biometrischen Fernerkennung, der Emotionserkennung mittels KI, oder einer KI, welche den Zugang zu Bildung ermöglicht oder verhindert: Es reicht einfach nicht aus, potenziell Betroffene auf die diesbezüglichen Gefahren von KI aufmerksam zu machen bzw. ausschließ auf die die Erhöhung der KI-Kompetenz zu setzen!

Daher muss im Sinne der Menschenzentrierung die Nutzung dieser und ähnlicher Use Cases auf Seiten der KI-Anbieter und Betreiber proaktiv so gewährleistet werden, dass die Souveränität des einzelnen nicht gefährdet werden darf. Der EU AI Act sieht einen entsprechenden Schutz an mehreren Stellen vor:

- Durch das in Artikel 5 EU AI Act formulierte Verbot manipulativer KI, des Social Scorings und anderer Use Cases wie z. B. der Fernidentifizierung in Echtzeit.
- In Artikel 6 (2) in Kombination mit Anhang III EU AI Act werden verschiedene Use Case als Hochrisiko-KI klassifiziert, z. B. solche zur Emotionserkennung oder im Hinblick auf die Zulassung zu Bildungseinrichtungen.
- Interessant ist auch die Regelung von Artikel 61 EU AI Act, der eine „informierte Einwilligung" aller Beteiligten erfordert, die an einem Test von Hochrisiko-KI unter Realbedingungen teilnehmen. Diese Teilnehmer besitzen das in Absatz 1 c) garantierte Recht, den Test ohne Begründung zu beenden (denkbar z. B. im Fall der Validierung von Medikamenten unter Verwendung von KI).

Die Souveränität des einzelnen, der in Alltag und Beruf mit einer für ihn kaum erkennbaren und noch weniger verständlichen KI-Welt konfrontiert werden kann, wird durch derlei spezifisch menschenzentrierte Vorgaben des EU AI Acts wirksam und angemessen geschützt.

Beschwerderecht

Der EU AI Act geht noch einen wichtigen Schritt weiter: Er gewährt in Artikel 85 EU AI Act jeder natürlichen Person das Recht zur Beschwerde bei einer Marktüberwachungsbehörde, wenn Grund zur Annahme hat, dass gegen die Bestimmungen der KI-Verordnung verstoßen wurde. Erweitert wird dies durch den expliziten Schutz von Hinweisgebern in Artikel 87 EU AI Act.

Der Druck zur Einhaltung aller in der KI-Verordnung auferlegten Pflichten für Anbieter und Betreiber von KI-Systemen als auch KI-Modellen sowie sonstiger Verpflichteter wird dadurch erheblich gesteigert: Beschwerden und Hinweise sind neben der Eigeninitiative von Behörden ein wichtiges Instrument, um im Fall der Fälle Verstöße aufzudecken und anschließend im Sinne von Artikel 99 ff. EU AI Act angemessen zu sanktionieren.

4. *Fazit und Ausblick*

Die vorherigen Ausführungen belegen, dass sich die in Artikel 1 EU AI Act gewährleistete Menschenzentrierung wie ein roter Faden durch die gesamte Norm zieht. Sie wird an vielen Stellen begleitend, Norm-übergreifend als auch spezifisch angemessen konkretisiert. Hinzu kommt die Bewehrung durch Sanktionen sowie ein Recht auf Beschwerde bzw. der Schutz von Hinweisgebern. Im Hinblick auf die Menschenzentrierung ist die KI-Verordnung definitiv ein „großer Wurf"!

In Anbetracht vielen der Möglichkeiten, die der EU AI Act bietet, um das menschenzentrierte Vertrauen in KI zu wahren bzw. zu erhöhen, darf ein Aspekt jedoch nicht zu kurz kommen: Dies ist die praxis- und zeitgerechte Interpretation. In dieser Hinsicht bleibt freiwilliges Engagement im Sinne der CDR nicht nur unverzichtbar – sie ist sogar noch wichtiger geworden: Künftig müssen stets aufs Neue Leitlinien erstellt, Best Practices gesammelt und bewertet werden, um den mit dem EU AI Act geschaffenen Rahmen angemessen auszufüllen.

Insofern wird auch in Zukunft das Zusammenspiel von CDR und EU AI Act von großer Bedeutung sein – nicht nur, aber insbesondere beim Thema Menschenzentrierung: Der EU AI Act, thematisiert, fördert und fordert sogar explizit an mehreren Stellen die Übernahme freiwilliger Verantwortung! So heißt es in Artikel 95 (1) EU AI Act: „Das Büro für Künstliche Intelligenz und die Mitgliedstaaten fördern und erleichtern die Aufstellung von Verhaltenskodizes, einschließlich damit zusammenhängender Governance-Mechanismen, mit denen die freiwillige Anwendung einiger oder aller der in Kapitel III Abschnitt 2 genannten Anforderungen auf KI-Systeme, die kein hohes Risiko bergen, gefördert werden soll."

Selbst wenn CDR an dieser und vergleichbaren Stellen nicht explizit erwähnt wird: Sie ist damit gemeint! In Anbetracht der exponentiellen Innovationsgeschwindigkeit wird CDR daher mit hoher Wahrscheinlichkeit ein wichtiges Instrument sein und bleiben, um das Vertrauen in KI zu bewahren bzw. zu steigern.

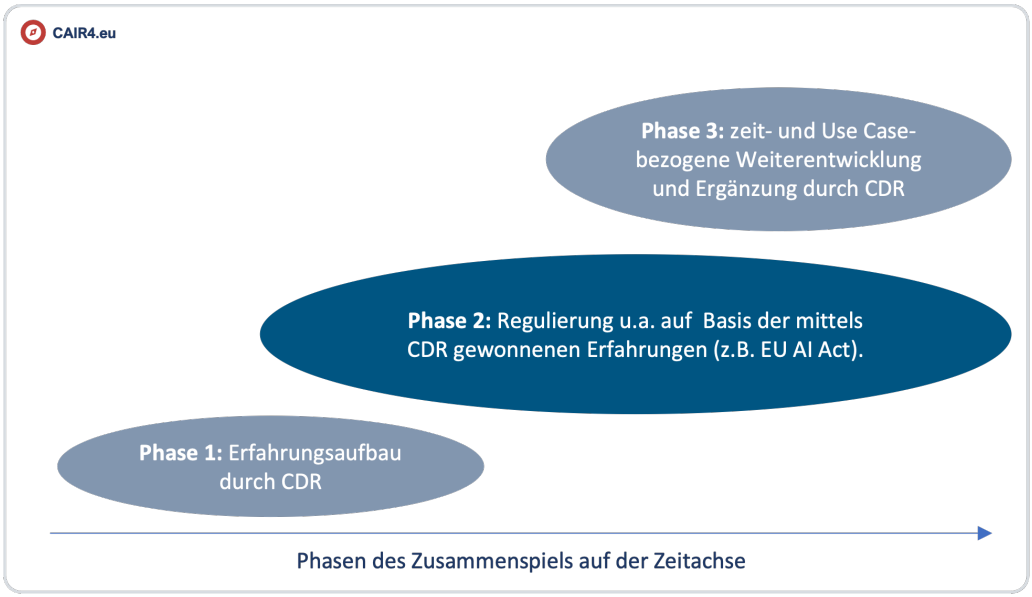

ABBILDUNG 5: ZUSAMMENSPIEL VON CDR UND KI-REGULIERUNG (QUELLE: EIGENE ABBILDUNG)

Insbesondere Deutschland hat aufgrund einer Vielzahl von Initiativen, Verbänden und Netzwerken die Möglichkeit, strukturiert, proaktiv und interdisziplinär zu handeln: Im Sinne einer „vierten Stufe" der unter Punkt 1.2 skizzierten Symbiose von CDR und Recht, also der laufenden Weiterentwicklung bereits bestehender Regulierung, die mutige und engagierte „First Mover" im Hinblick auf die freiwillige Übernahme digitaler Verantwortung belohnt – und nicht nur die „Sanktions-Peitsche" schwingt.

So gesehen sind Instrumente wie die CDR-Initiative des BMUV sowie CDR-Frameworks als auch der CDR-Award in Zukunft noch wichtiger als bisher. Im Hinblick auf menschenzentrierte KI sollten sie allerdings noch mehr mittlerweile erfolgreich gesteckten regulativen Rahmen aufgreifen und gezielt weiterentwickeln!

Accenture (2020): Adding a Human-Centered Approach to Business, URL: https://www.accenture.com/us-en/insights/strategy/human-centered-business (aufgerufen am: 01/07/2024).

BDI (2024): Human Centricity, URL: https://bdi.eu/spezial/forward-to-the-new/human-centricity#/artikelwide/news/1 (aufgerufen am: 01/07/2024).

Breinrich-Schilly, A. (2023): KI und Nachhaltigkeit spielen für Investoren zentrale Rolle, URL: https://www.springerprofessional.de/investition/csr-reporting/investoren-setzen-auf-ki-in-portfoliounternehmen/26327062 (aufgerufen am: 05/09/2024).

Bundesregierung (2018): Nationale KI-Strategie, URL: https://www.ki-strategie-deutschland.de/ (aufgerufen am: 05/09/2024).

Carli, R. / Najjar, R. / Calvaresi, D. (2022): Risk and Exposure of XAI in Persuasion and Argumentation: The Case of Manipulation, in: Clavaresi, D. / Najjar, A. / Winikoff, M. / Främling, K. (Eds.): Explainable and Transparent AI and Multi-Agent Systems, Heidelberg, New York: Springer, 204–220.

CDR (2024): CDR-Manifesto, URL: https://corporatedigitalresponsibility.net/cdr-manifesto (aufgerufen am: 05/09/2024).

DIN (Hrsg.) (2019): Ergonomie der Mensch-System-Interaktion – Teil 210: Menschzentrierte Gestaltung interaktiver Systeme (ISO 9241-210:2019).

Europäische Kommission (2022): Cyber Resilience Act, URL: https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act (aufgerufen am: 05/09/2024).

Europäisches Parlament / Rat der Europäischen Union (2017): Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates, URL: https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32017R0745 (aufgerufen am: 05/09/2024).

– (2024): Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates, URL: https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L_202401689 (aufgerufen am: 05/09/2024).

Eviden (2024): Tech Trend Radars, URL: https://eviden.com/publications/tech-radar/artificial-intelligence/ (aufgerufen am: 05/09/2024).

Fyler, T. (2023): The Big Red Button: Why We Need an AI Kill Switch, URL: https://techhq.com/2023/08/will-a-big-red-button-add-security-to-generative-ai/ (aufgerufen am: 05/09/2024).

Identity Valley (2024): Digital Responsibility Goals, URL: https://identityvalley.eu/drg 2024 (aufgerufen am: 05/09/2024).

Merx, O. (2024): KI-Kompetenz: Pflicht oder freiwillig?, URL: https://cair4.eu/ki-kompetenz-pflicht-oder-freiwillig (aufgerufen am: 28/10/2024).

Mütze, S. (2022): Corporate Social Responsibility and the Effects of Sustainable Corporate Practices and Various Greenwashing Methods on Corporate Reputation, in: Junior Management Science, Vol. 7 / No. 3, 826–873.

Phoenix (2018): Datenskandal. Facebook-Datenaffäre, URL: https://www.phoenix.de/datenskandal-a-142427.html (aufgerufen am: 05/09/2024).

Whyzer (2021): Vorstellung und Vergleich: Der CDR-Kodex und die CDR Building Bloxx, URL: https://www.whyzer.io/vergleich-bmjv-cdr-kodex-und-bvdw-building-bloxx (aufgerufen am: 05/09/2024).