1.1.5 Kriterien zur Erklärbarkeit der 4.0-Technologien

Stichwörter: Algorithmen, Datensouveränität, Deep Learning, künstliche Intelligenz, Sensorik, Transparenz

› Warum ist das Thema wichtig?

Cyber-physische Systeme[1] und intelligente Software[2] mit ihren Modellen der künstlichen Intelligenz (KI) erfassen im Arbeitsprozess Daten, verarbeiten sie, lernen und geben Daten weiter (Big Data). Dies geschieht oft, ohne das Betreiber und Nutzer genau wissen, welche Daten erfasst werden und nach welchen Kriterien beziehungsweise Regeln die Software 4.0 (inkl. KI) lernt oder Entscheidungen trifft oder wie die Daten verwendet werden. Die Anwender/Nutzer verfügen nicht mehr allein über ihre Daten (Datensouveränität). Die Potenziale und Möglichkeiten der 4.0-Technologien[3] für den Wertschöpfungsprozess sowie für eine sichere und gesundheitsgerechte Arbeitsgestaltung können jedoch dann ausgeschöpft werden, wenn bekannt ist, welche Daten erfasst und anhand welcher dahinterliegenden Entscheidungsregeln diese miteinander in Verbindung gebracht werden. Aus diesen Gründen ist es wichtig, dass die Betreiber und die Nutzer der Daten grundsätzlich darüber informiert sind, welche Daten in den 4.0-Prozessen[4] erhoben werden, wie die intelligente Software (inkl. KI) arbeitet und die Abläufe ganz oder teilweise steuert, wie sie lernt, wo die Daten wie verwendet und wo sie gespeichert werden.

› Worum geht es bei dem Thema?

Begriffe: Transparenz – Erklärbarkeit
Unter **Datensouveränität**[5] wird hier verstanden, dass ein Nutzer von 4.0-Technologien (zum Beispiel Betreiber, Führungskraft, Beschäftigter) weiß, was mit seinen erfassten Daten geschieht, und der Verwendung zustimmt (inklusive Datenschutz und Datensicherheit). Datensouveränität ist die „Fähigkeit einer natürlichen oder juristischen Person zur ausschließlichen Selbstbestimmung hinsichtlich des Wirtschaftsguts Daten“.[6] Grundlage ist die Befugnis und die Entscheidungskompetenz, selbst zu bestimmen, mit welchen Inhalten jemand in Beziehung zu seiner Umwelt tritt und mit ihr kommuniziert. Gleichzeitig wird mit der Zustimmung die Möglichkeit der Nutzung der Daten in Big Data und 4.0-Prozessen möglich. Der Umfang der Datensouveränität ist abhängig vom Umfang der Transparenz und Erklärbarkeit der 4.0-Prozesse.
Transparenz bezeichnet bei 4.0-Technologien die Eigenschaft, dass die Aktionen und Funktionen des autonomen technischen Systems nachvollziehbar sind. Die Forderung nach maximaler Transparenz ist häufig nicht vollständig erfüllbar, da viele Modelle so komplex sind, dass Nutzer von 4.0-Technologien diese softwaretechnischen Abläufe nicht durchschauen können.[7]
Erklärbarkeit: Damit 4.0-Technologien erklärbar sind, müssen die Gesetzmäßigkeiten, nach denen das autonome technische System agiert, bekannt sein.

In den Umsetzungshilfen „Arbeit 4.0“ wird immer wieder die Transparenz und Erklärbarkeit der 4.0-Technologien als Maßnahme gefordert, da Betreiber und Nutzer wissen sollten, welche Daten erhoben werden, wie sie verarbeitet werden, nach welchen Kriterien die intelligente Software (inkl. KI) Entscheidungen trifft und die Prozesse ganz oder teilweise steuert, sowie nach welchen Kriterien die intelligente Software lernt und wie mit den Daten des Betriebes beziehungsweise der beteiligten Personen umgegangen wird. Das ist die Voraussetzung, um die Daten (des Betriebes, der Personen) bewusst, gezielt und systematisch für die Wertschöpfungsprozesse nutzen zu können sowie Datenschutz und Datensicherheit der eigenen Daten und der genutzten intelligenten Software (inkl. KI) zu realisieren.

Betriebe und Personen sollten ein Bewusstsein herausbilden, wie 4.0-Technologien arbeiten und wie ihre Daten verwendet werden. Dazu ist kein spezifisches IT-Wissen erforderlich. Die wesentlichen Kriterien des Umgangs mit den betrieblichen und personenbezogenen Daten sollten sie jedoch kennen.

Im Folgenden werden Kriterien für die Erklärbarkeit der 4.0-Technologien als Grundlage für die Datensouveränität beschrieben, über die Betriebe und Personen vor dem Einsatz von 4.0-Tech-

Diese Umsetzungshilfe gibt Experten und Interessierten Anregungen, wie Arbeit 4.0 zu gestalten ist. Die Empfehlungen sollten an die jeweilige konkrete betriebliche Situation angepasst werden.

[1] Cyber-physische Systeme (CPS) verbinden und steuern als autonome technische Systeme Arbeitsmittel, Produkte, Räume, Prozesse und Menschen beinahe in Echtzeit. Die komplette oder teilweise Steuerung übernimmt intelligente Software auf Grundlage von Modellen der künstlichen Intelligenz. Genutzt werden dazu unter anderem auch Sensoren/Aktoren, Verwaltungsschalen, Plattformen/Clouds.

[2] Intelligente Software steuert cyber-physische Systeme (CPS) und andere autonome technische Systeme (wie Messenger-Programme). Intelligente Software nutzt Modelle künstlicher Intelligenz zusammen mit anderen Basistechnologien wie zum Beispiel Algorithmen, semantischen Technologien, Data-Mining. Intelligente Software ist autonom und selbstlernend.

[3] 4.0-Technologie bezeichnet hier Hardware und technologische Produkte (wie Assistenzmittel/Smartphones, Sensoren/Aktoren in smarten Arbeitsmitteln, Fahrzeugen, Produkten, Räumen usw., smarte Dienstleistungen, Apps), die von intelligenter Software (inkl. KI) ganz oder teilweise gesteuert werden.

[4] Unter 4.0-Prozessen werden hier alle Arbeitsprozesse verstanden, in denen cyber-physische Systeme (CPS) oder andere autonome technische Systeme (wie Plattformen, Messenger-Programme) beteiligt sind. 4.0-Prozesse sind in den Arbeitsprozessen bisher selten vollständig, aber in Ansätzen in allen Betrieben umgesetzt.

[5] vgl. Deutscher Ethikrat 2017, S. 166ff.; Gräf et al. 2018, S. 4ff.; Otto 2016

[6] Otto 2016, S. 5

[7] vgl. BaFin 2018, S. 37; Döbel et al. 2018, S. 30

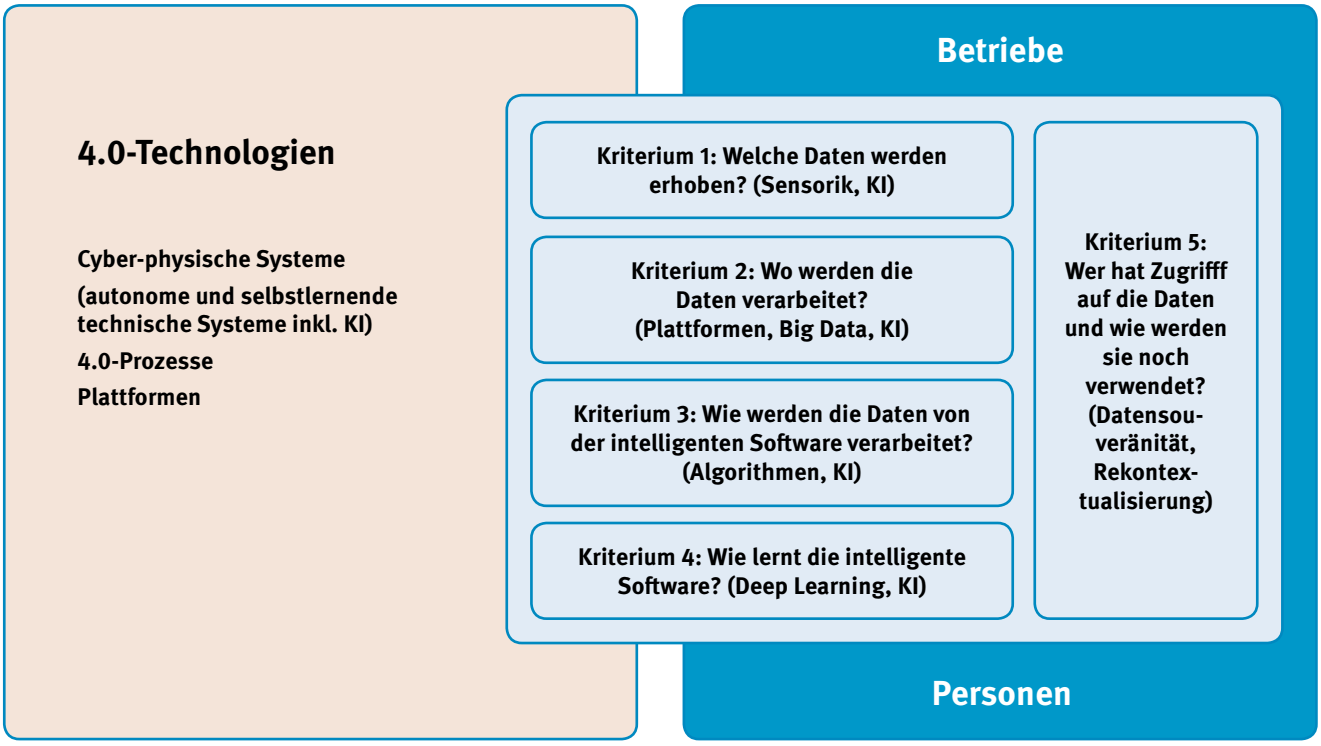

Abbildung 1: Kriterien für die Erklärbarkeit der 4.0-Technologien (eigene Darstellung)

nologien Informationen von Herstellern, Anbietern, Dienstleistern und Programmierern einfordern können – *siehe Abbildung 1*.

Die Kriterien für die Erklärbarkeit der 4.0-Technologien werden im Folgenden weiter erläutert:

Kriterium 1: Betreiber und Nutzer sollten wissen, welche Sensoren in dem „Gegenstand" sind, den sie anschaffen oder benutzen wollen, und welche Daten von diesen Sensoren erhoben werden.

Sensoren sind (elektronische) Bauteile zur Erfassung einer physikalischen Größe. Sensoren können alle physikalisch messbaren Werte erfassen[8], wie zum Beispiel Wärmestrahlung, Temperaturen, Feuchtigkeit, Druck, Kraft, Beschleunigung oder Magnetismus. Damit können Daten über Zustände, Bewegungen und Prozesse von Arbeitsmitteln, Fahrzeugen, Arbeitsstoffen, Räumen, Assistenzmitteln, Produkten über integrierte Sensorik erfasst werden. Aber auch personenbezogene Daten (wie zum Beispiel Vitaldaten, Bewegungsdaten) können über Sensorik beispielsweise in Smartphones, Kleidung oder Fitnessarmbändern erhoben werden.

Sensoren können mikroskopisch klein sein, sodass sie beispielsweise in Gewebefäden von Kleidung integriert werden. Es existieren bereits winzige Sensoren („Smart Dust" – intelligenter Staub). Viele Sensorsysteme basieren auf der Kombination sehr unterschiedlicher Sensortypen. Fitnessarmbänder enthalten beispielsweise Bewegungssensoren, optische Sensoren, bioelektrische Sensoren und GPS-Empfänger. Die Informationen dieser Sensorsysteme können sich ergänzen oder gegenseitig kontrollieren (Sensorfusion).

Um Daten für seine Geschäftszwecke und seine Arbeitsgestaltung zu erhalten und um Datenschutz und Datensicherheit garantieren zu können, sollte jeder Betrieb vom Hersteller/Dienstleister Informationen darüber einfordern, welche wesentlichen Sensoren in dem smarten Gegenstand sind und welche Daten diese erheben (betriebliche und personenbezogene).

Kriterium 2: Betreiber und Nutzer sollten wissen, wo Daten gespeichert werden.

Die Daten, die über Sensoren erfasst werden oder die selbst eingegeben werden (zum Beispiel über Messenger-Dienste, Social Media, Suchmaschinen, Internetnutzung, E-Mails), werden auf Plattformen abgelegt › *Siehe Umsetzungshilfe 2.5.3 Plattformökonomie*. So werden beispielsweise die Daten des Fahrzeugs in der Regel auf einer Plattform des Herstellers gesammelt, die Daten des Fitnessarmbands auf einer Service-Plattform oder die in eine Suchmaschine eingegebenen Daten auf der Plattform des Betreibers der Suchmaschine. In der Regel wissen Betreiber und Nutzer nicht, wo ihre Daten gespeichert werden. In den Allgemeinen Geschäftsbedingungen (AGB) der Anbieter und Hersteller sollte diese Information zu finden sein. Diese AGB sind meist sehr umfassend. Kurzinformationen dazu stehen oft nicht zur Verfügung.

Betreiber und Nutzer sollten sich dennoch informieren, wo ihre Daten gespeichert werden, bevor sie den smarten „Gegenstand" oder die Anwendung anschaffen. Auch bei bereits verwendeter 4.0-Technologie sollte zum Beispiel Fragen der Datensicherheit und des Datenschutzes nachgegangen werden. Relevant kann auch der Gerichtsort sein, da das Recht des Landes angewendet wird, in dem die Daten liegen.

[8] vgl. Cernavin & Lemme 2018, S. 27ff.; Mücklich 2015; Völz 1999

Kriterium 3: Betreiber und Nutzer sollten wissen, nach welchen Kriterien die intelligente Software (inkl. KI) ihre Daten verarbeitet und nach welchen Kriterien sie die Prozesse steuert (teilweise oder komplett).

Intelligente Software (inkl. KI) verarbeitet, entscheidet und steuert (ganz oder teilweise) Prozesse mit ihren Modellen der KI ganz oder teilweise, ohne dass dies von außen sichtbar ist. Ebenso wenig können Betreiber und Nutzer oftmals erkennen, warum die intelligente Software (inkl. KI) zu welchem Ergebnis kommt. Um dennoch zu wissen, nach welchen Kriterien die intelligente Software (inkl. KI) Daten verarbeitet und Prozesse ganz oder teilweise steuert, lassen sich folgende Empfehlungen ableiten:[9]

Betreiber oder Nutzer sollten beim Einsatz von intelligenter Software (inkl. KI) vom Hersteller/Dienstleister zum Beispiel folgende Informationen beschaffen:

- *Grundlegende Zielsetzung:* Die Zielsetzung der intelligenten Software (inkl. KI) sollte allgemeinverständlich dargestellt sein. Dabei geht es nicht um die softwaretechnische Zielsetzung, sondern um die inhaltlichen Vorgaben, die die intelligente Software (inkl. KI) erfüllen soll (zum Beispiel Konzepte für den sicheren Umgang mit einer Maschine, Optimierung von Persönlichkeitsprofilen zur Verbesserung der Personaleinsatzplanung).
- *Inhaltliche Entscheidungskriterien:* Die Entscheidungsfindung der intelligenten Software (inkl. KI) sollte für einen Laien nachvollziehbar sein (Handlungslogik). Dabei geht es nicht um die Offenlegung von Quellcodes, Algorithmen oder Programmstrukturen, sondern darum zu verstehen, auf Grundlage welcher inhaltlichen Kriterien und Gewichtungen die Software entscheidet.[10] Beispiele hierzu sind:
  - Bei der Steuerungssoftware von smarten verketteten Arbeitsmitteln sollte dargestellt sein, nach welchen Kriterien sie die Prozesse ganz oder teilweise steuert, wie zum Beispiel Zeit, Qualität, Materialeinsatz, Verfügbarkeit von Personal › *Siehe Umsetzungshilfe 3.1.4 Sicherheit von verketteten Arbeitsmitteln mit 4.0-Technologie.*
  - Bei der smarten Personaleinsatzplanung sollte informiert werden, nach welchen Kriterien sie Personal aussucht, wie zum Beispiel Kompetenz, gerechte Arbeitszeitverteilung, Gesundheitszustand › *Siehe Umsetzungshilfe 2.6.1 Digitale Planung des Personaleinsatzes.*
  - Die Personalbewertungssoftware sollte die Kriterien darstellen, nach denen die Personen bewertet werden und wie diese gewichtet werden, zum Beispiel Fehlzeiten, Arbeitsproduktivität, Arbeitszeit, Fehltage, Zufriedenheit › *Siehe Umsetzungshilfe 2.6.3 Personalbeurteilung und CPS.*

Es ist auch möglich, dass Daten nicht für den vorgesehenen Zweck genutzt, sondern ohne Wissen des Nutzers interpretiert und zusätzliche Schlussfolgerungen daraus gezogen werden. Zu einer derartigen Datenanalyse, die auch diskriminierend[11] sein kann, kann es beispielsweise kommen, wenn die intelligente Software (inkl. KI) ein geändertes (Einkaufs-)Verhalten erkennt und daraus auf bevorstehende Lebensereignisse schließt (wie Trennungen, Schwangerschaften), ohne dass hierzu konkrete Daten erhoben worden sind. Über derartige Möglichkeiten der intelligenten Software (inkl. KI) sollte informiert werden.

- *Einbindung von Plattformen:* Der Zugriff von intelligenter Software (inkl. KI) auf Plattformen von weiteren Anbietern durch den Hersteller/Dienstleister sollte allgemeinverständlich dargestellt sein. Auch hier geht es nicht um die softwaretechnischen Codes, sondern um die Benennung der Anbieter, auf deren Daten und Plattformen zurückgegriffen wird (zum Beispiel Social Media, Hersteller).
- *Sicherheit und Gesundheit bei der Arbeit:* Regelungen und Grundlagen des sicheren und gesundheitsgerechten Arbeitens, die die intelligente Software (inkl. KI) bei ihren Entscheidungen berücksichtigt, sollten durch den Hersteller/Dienstleister allgemeinverständlich dargestellt werden (zum Beispiel Verordnungen, DGUV Vorschriften, technische Regeln, Normen und arbeitswissenschaftliche Erkenntnisse).
- *Berücksichtigung weiterer ethischer Kriterien:* Die ethischen Grundlagen, die die intelligente Software (inkl. KI) berücksichtigt, sollten bekannt sein, wie zum Beispiel Fairness und Diskriminierungsfreiheit.[12] › *Siehe Umsetzungshilfe 1.1.4 Ethische Werte für die intelligente Software.*
- *Datenschutz und Datensicherheit:* Es sollte bekannt sein, wie die intelligente Software (inkl. KI) bei ihren Entscheidungen Datenschutz (zum Beispiel Datenschutz über Privacy by Design[13], Privacy-Preserving Data-Mining[14]) und Datensicherheit (zum Beispiel über Zertifikate wie Trusted Cloud) sicherstellt[15] › *Siehe Umsetzungshilfen 2.3.1 Datensicherheit in 4.0-Prozessen; 2.3.2 Datenschutz in 4.0-Prozessen.*

Diese Empfehlungen sind sowohl auf kommerzielle Software als auch auf Open-Source-Software (offener Quellcode) anwendbar. › *Siehe Umsetzungshilfe 2.1.2 Integration von intelligenter Software in die Organisation.*

Kriterium 4: Betreiber und Nutzer sollten wissen, nach welchen Kriterien die intelligente Software (inkl. KI) lernt und sich autonom weiterentwickelt.

Die Kriterien, nach denen intelligente Software autonom lernt und sich weiterentwickelt (Lernalgorithmen, künstliche Intelligenz, Deep Learning), sind für den Anwender in vielerlei Hinsicht unverständlich und nicht direkt erkennbar.[16]

Bei lernender intelligenter Software (inkl. KI) und insbesondere beim Deep Learning[17] werden beispielsweise Lernalgorithmen sowie ständig wachsende Datenmengen in große künstliche Lernnetzwerke eingespeist. Diese Lernnetz-

[9] vgl. u. a. Bünau 2018; Bundesregierung 2016, S. 5f.; Busch 2018, S. 40ff.; Deutscher Ethikrat 2017, S. 166ff.; Gräf et al. 2018; Krüger & Lischka 2018, S. 31ff.; Rohde 2018
[10] Wachter 2018
[11] Kar et al. 2004, S. 478
[12] Döbel et al. 2018, S. 31
[13] Privacy by Design = Die Software 4.0 verarbeitet prinzipiell keine Informationen, die in Bezug auf den Datenschutz problematisch sind – siehe BaFin 2018, S. 38
[14] Privacy-Preserving Data-Mining = Die Software 4.0 integriert die Datenschutzanforderungen direkt in die Datenanalyse und Datenauswertung – siehe BaFin 2018, S. 38
[15] Döbel et al. 2018, S. 31
[16] vgl. u. a. BaFin 2018 S. 37; Stockley 2018
[17] „Deep" bezieht sich auf die vielen Schichten, die das künstliche Lernnetzwerk mit der Zeit autonom ansammelt. Je vielschichtiger das Netzwerk wird, desto stärker werden die Leistung und die Lernfähigkeit im Rahmen der vorgegebenen Parameter.

werke orientieren sich an der Funktionsweise von neuronalen Netzwerken des menschlichen Gehirns. Die Fähigkeit der Lernnetzwerke zu „denken“ und zu „lernen“ steigt, je mehr Daten verarbeitet werden. Die künstlichen Lernnetzwerke der intelligenten Software (inkl. KI) sollten mit einem spezifischen Lernziel von Auftraggebern programmiert werden und sie lernen dann autonom weiter. Problem bei diesen Lernprozessen kann sein, dass der lernenden Software Daten zugrunde liegen, die für die Fragestellung des Betriebes nicht zutreffen, oder dass die Qualität der Daten nicht ausreicht. › *Siehe Umsetzungshilfe 2.3.3 Datenqualität in 4.0-Prozessen.* Es kann auch sein, dass die lernende Software nicht zulässige oder nicht sinnvolle Korrelationen herstellt und nicht korrekte Schlussfolgerungen zieht. Zudem ist es möglich, dass intelligente Software (inkl. KI) ohne Lernziel ausgestattet ist und autonom in eine gegebenenfalls nicht gewünschte Richtung weiterlernt.

Damit kleine und mittlere Betriebe, Führungskräfte und Beschäftigte annähernd erkennen können, nach welchen Kriterien die intelligente Software (inkl. KI) lernt und sich weiterentwickelt, können unter anderem folgende Empfehlungen hilfreich sein:[18]

- *Inhaltliches Lernziel:* Es sollte bekannt sein, welches Lernziel die intelligente Software (inkl. KI) besitzt. Dabei geht es um inhaltliche Ziele, die vorgegeben werden (zum Beispiel dass ein autonomes technisches System die Maschine ergonomisch optimal auf den Benutzer einstellen soll).
- *Inhaltliche Lernkriterien:* Es sollte bekannt sein, nach welchen inhaltlichen Kriterien die intelligente Software (inkl. KI) lernt (Lernlogik). Hier geht es darum aufzuzeigen, nach welchen allgemeinen inhaltlichen Parametern die intelligente Software (inkl. KI) auf Grundlage des Lernziels lernt. Inhaltliche Parameter können zum Beispiel sein: Verkettung von Arbeitsmitteln zur Optimierung der Schnittstellen Mensch – Maschine, die wirkungsvollere Ausnutzung der einzelnen Maschinen zur Optimierung der Produktion.
- *Datengrundlagen:* Es sollte bekannt sein, auf welcher allgemeinen inhaltlichen Datengrundlage die intelligente Software (inkl. KI) lernen kann (ihre Lernmuster herausbilden kann). Beispielsweise kann bekannt sein, dass eine Software, die Unfallursachen analysiert und Unterweisungshinweise gibt, auf Grundlage der Unfallberichte einer Branche ihre Lernmuster entwickelt hat.
- *Ethische Grundlagen für das Lernen:* Es sollte bekannt sein, ob beziehungsweise wie die intelligente Software (inkl. KI) beim Lernen ethische Grundlagen berücksichtigen kann, wie zum Beispiel Sicherheit und Gesundheit oder Fairness und Diskriminierungsfreiheit.[19] › *Siehe Umsetzungshilfe 1.1.4 Ethische Werte für die intelligente Software.*
- *Datenschutz und Datensicherheit:* Es sollte bekannt sein, ob beziehungsweise wie die intelligente Software (inkl. KI) im Lernprozess den Datenschutz und die Datensicherheit (zum Beispiel über Zertifikate wie Trusted Cloud) sicherstellt. › *Siehe Umsetzungshilfen 2.3.1 Datensicherheit in 4.0-Prozessen; 2.3.2 Datenschutz in 4.0-Prozessen.*

Kriterium 5: Betreiber und Nutzer sollten wissen, wer auf ihre Daten Zugriff hat sowie ob und wie diese noch verwendet werden.

Betreiber, Führungskräfte und Beschäftigte sollten wissen, wer Zugriff auf ihre Daten hat und in welchen Zusammenhängen ihre Daten verwendet werden. Da die Daten des Betriebes oder der Person, mit der die intelligente Software (inkl. KI) arbeitet, oft auf Plattformen liegen oder in anderen betriebsübergreifenden Zusammenhängen verwendet werden, die nicht nur dem Betreiber oder Nutzer zugänglich sind, können die Daten noch von weiteren Anwendern genutzt werden. › *Siehe Umsetzungshilfe 2.5.3 Plattformökonomie.* Nutzer sollten sich darüber informieren, was mit den anonymisierten Daten passiert. Beispielsweise sollte Nutzern bekannt sein, dass die anonymisierten Daten eines Fahrzeugs zur Optimierung der Fahrleistung und zur optimalen Einstellung der Fahrzeugtechnik verwendet werden. Werden diese Daten an Dritte weitergegeben, sollte auch darüber informiert werden. Beispielsweise werden die erfassten Daten mit den Daten anderer Hersteller abgeglichen und von dem Institut XY für statistische Zwecke ausgewertet.

4.0-Technologien ermöglichen darüber hinaus eine umfassende Dekontextualisierung und Rekontextualisierung von Daten. Das bedeutet, die Daten werden für Zusammenhänge genutzt, für die sie nicht erhoben wurden, und sie werden neu verknüpft.[20] Dies kann bei betrieblichen und personenbezogenen Daten sehr sinnvoll, aber auch problematisch sein, wenn beispielsweise unbefugt Nutzerprofile erstellt werden. Diese Vorgänge ermöglichen zum Beispiel im Gesundheitswesen, dass in der Klinik erhobene Laborwerte mit in Forschungslaboren durchgeführten Gesamtgenomanalysen verknüpft werden und so verbesserte Therapiemöglichkeiten für individuelle Patienten ausgewählt werden können.[21] Dies ist zum einen aus Datenschutzgründen ohne Zustimmung des Betroffenen grundsätzlich nicht zulässig. Zum anderen können die Daten für das Profiling ungeeignet sein beziehungsweise können sie für Fragestellungen (zum Beispiel Verhaltensmuster) genutzt werden, für die sie nicht erhoben wurden (zum Beispiel kann das Kaufverhalten einer Person Rückschlüsse auf ihre Gesundheitssituation zulassen). Dabei kann es zu Verzerrungen kommen, wenn Daten nicht lückenlos vorliegen (zum Beispiel Daten über Sportaktivitäten, die nicht erfasst werden, oder fehlende Gesundheitsdaten).

Auch in der Frage des Zugriffs auf die Daten sollten Betriebe und Personen vom Hersteller/Dienstleister allgemeine grundlegende inhaltliche Informationen einfordern, wie mit ihren Daten umgegangen wird. Dazu gehört, wer Zugriff auf die Daten hat und in welchen Zusammenhängen sie von wem verwendet werden.

[18] vgl. u. a. Anderl 2018; BaFin 2018, S. 26ff.; Deutscher Ethikrat 2017, S. 166ff.; Döbel et al. 2018; Kar et al. 2017; Rohde 2018; Schonschek 2018; Stockley 2018; Wachter 2018
[19] Döbel et al. 2018, S. 31
[20] Deutscher Ethikrat 2017, S. 57
[21] Deutscher Ethikrat 2017, S. 28
[22] Busch 2018, S. 59

Welche Chancen und Gefahren gibt es?

Wenn Datensouveränität gewährleistet ist, bietet dies unter anderem folgende **Chancen**:

- Kontrolle über die Steuerung der Lernprozesse von intelligenter Software (inkl. KI).
- Gezielte Einflussnahme auf das Lernziel der intelligenten Software (inkl. KI) während des Lernprozesses.
- Systematische Datenerfassung durch Sensorik kann nützlich sein (beispielsweise Erfassung von Vitaldaten im Gesundheitswesen).
- Speicherort der Daten ist bekannt und der Zugriff darauf geregelt.
- Optimierung von Prozessen möglich (zum Beispiel effizientes Fahren).
- Frühzeitiges und vorausschauendes Erkennen von gesundheitlichen oder technischen Problemen.
- Höheres Vertrauen in intelligente Software (inkl. KI) und bessere Akzeptanz.
- Bessere Nutzbarkeit von intelligenter Software (inkl. KI), höhere Produktivität und Zufriedenheit der Beschäftigten, wenn Kenntnisse über deren Funktionen vorliegen und sich ein Bewusstsein im Umgang damit entwickelt.
- Sicherstellung der Betriebssicherheit.

Wenn Datensouveränität nicht geklärt ist, ergeben sich daraus unter anderem folgende **Gefahren**:

- Kontrollverlust über das Lernziel der intelligenten Software (inkl. KI).
- Nutzer kennen die Kriterien nicht, nach denen intelligente Software (inkl. KI) lernt und sich weiterentwickelt.
- Bekanntgabe von Daten, die nicht bekannt sein sollen und damit beispielsweise unbefugtes Erstellen von Nutzerprofilen.
- (Unbekannte) Sensoren erfassen Daten, von denen Nutzer nichts wissen.
- Daten können im Ausland gespeichert werden und den dortigen rechtlichen Bestimmungen unterliegen.
- Unbefugter Zugriff zu den Daten und unbefugte Nutzung durch Dritte.
- Gewährleistung der Betriebssicherheit nicht möglich, wenn diese Kriterien im Lernziel nicht definiert sind.
- Misstrauen und fehlende Akzeptanz der Betroffenen in Bezug auf die intelligente Software (inkl. KI).
- Mögliche Konflikte zwischen Unternehmer, Führungskräften und Beschäftigten beziehungsweise zwischen Betriebsrat und Geschäftsführung mit gegebenenfalls unnötigen Vorbehalten gegenüber der 4.0-Technologie generell.
- Ethische Grundlagen sind nicht definiert und mit Betreibern und Nutzern nicht kommuniziert.

Welche Maßnahmen sind zu empfehlen?

Bei allen Anwendungen von 4.0-Technologien – vom einfachen smarten Handbohrer über das Smartphone und das Fahrzeug bis hin zu komplexen autonom gesteuerten Prozessen – sollten Betreiber und Nutzer vom Hersteller/Dienstleister dieser Systeme allgemeinverständliche Erklärungen zur Funktionsweise der autonomen technischen Systeme einfordern. Dies sollte in jedem Fall bei Neuanschaffungen oder Programmierungen der Fall sein, es sollten aber gegebenenfalls auch bestehende 4.0-Prozese und -Anwendungen hinterfragt werden.

Dabei sollten Betreiber und Nutzer unter anderem folgende Aspekte hinterfragen beziehungsweise dazu Informationen einfordern:

Welche Daten werden erhoben (Sensorik)?

- Welche Sensoren sind in der 4.0-Technologie integriert?
- Welche Daten erheben diese Sensoren?

Wo werden die Daten gespeichert (Plattformen)?

- Auf welchen Plattformen werden die Daten gespeichert?
- Welches Landesrecht gilt?
- Wie ist die Datensicherheit der Plattform geregelt (zum Beispiel zertifiziert durch Trusted Cloud)?

Wie werden die Daten von der intelligenten Software verarbeitet (Algorithmen, künstliche Intelligenz)?

- Welche grundlegende Zielsetzung hat die intelligente Software (inkl. KI)?
- Nach welchen allgemeinen inhaltlichen Kriterien entscheidet die intelligente Software (inkl. KI)?
- Werden auch Merkmale erarbeitet, die nicht speziell erfasst werden (entstehen zum Beispiel durch Kombination erhobener Merkmale neue Merkmale)?
- Greift die intelligente Software (inkl. KI) auf Daten von anderen Plattformen zurück, wie verwertet die Software 4.0 diese Daten und welche Schlussfolgerungen zieht sie daraus?
- Wie berücksichtigt die intelligente Software (inkl. KI) das Thema Sicherheit und Gesundheit bei der Arbeit?
- Wie werden im Lernprozess der intelligenten Software (inkl. KI) und bei autonomen Entscheidungen Datenschutz und Datensicherheit berücksichtigt?

Wie lernt die intelligente Software (wie Deep Learning, KI)?

- Welches Lernziel verfolgt die intelligente Software (inkl. KI)?
- Nach welchen allgemeinen inhaltlichen Kriterien lernt die intelligente Software(inkl. KI)?
- Auf welcher allgemeinen inhaltlichen Datengrundlage hat die intelligente Software (inkl. KI) ihre Lernmuster erstellt?
- Auf welcher allgemeinen inhaltlichen Datengrundlage lernt die intelligente Software (inkl. KI) weiter (zum Beispiel inhaltliche Kriterien des Lernens wie Anforderungen aus Gesetzen, ökonomische Kriterien)?
- Wie berücksichtigt die intelligente Software (inkl. KI) ethische Grundlagen beim Lernen (wie zum Beispiel Sicherheit und Gesundheit des Menschen oder Fairness und Diskriminierungsfreiheit)?
- Wie berücksichtigt die intelligente Software (inkl. KI) Datenschutz und Datensicherheit beim Lernen?

Wer hat Zugriff auf die Daten und wie werden sie noch verwendet (Datensouveränität)?

- Wer hat Zugriff auf die Daten von Betreibern/Nutzern?
- In welchen Zusammenhängen werden Daten von Betreibern/Nutzern über die konkrete Nutzung durch Betreiber/Nutzer hinaus zusätzlich verwendet?

- Werden die Daten von Betreibern/Nutzern für Zusammenhänge genutzt, für die sie nicht erhoben wurden (Dekontextualisierung), und werden sie neu verknüpft (Rekontextualisierung)?

Die Darstellung dieser Informationen ließe sich dabei sehr gut in kompakter Form realisieren:

Empfehlung: Kurzinformation zum Umgang mit Daten
Die Kriterien zur Erklärung des Umgangs mit den Daten von einer autonomen und selbstlernenden Software (inkl. KI) sollten für Laien so aufbereitet werden, dass sie schnell nachzulesen und verständlich sind. Sie sollten kurz und übersichtlich formuliert sein und die wesentlichen Informationen, zum Beispiel zur Erfassung, Speicherung, Weitergabe von Daten, enthalten (zum Beispiel in Form eines „Onepagers“).[22]
Der Verweis auf oftmals ausführliche und komplex formulierte Allgemeine Geschäftsbedingungen (AGB) hilft im Alltag von kleinen und mittleren Unternehmen oft nicht weiter.
Beispielsweise werden vom Hersteller nach Gefahrstoffverordnung Sicherheitsdatenblätter über Gefahrstoffe zur Verfügung gestellt, zu denen der Arbeitgeber Kurzinformationen erstellt, die als Betriebsanweisung bekannt sind.
› *Siehe Umsetzungshilfe 1.1.7 Informationsblatt smartes Produkt.*

Quellen und weitere Informationsmöglichkeiten:

Anderl, S. (2018). *KÜNSTLICHE INTELLIGENZ: Denn wir wissen nicht, wie sie's tun.*http://www.FAZ.NET/AK1.1.6 Informationsblatt smartes ProduktTUELL/WISSEN/COMPUTER-MATHEMATIK/DIE-RISIKEN-KUENSTLICHER-INTELLIGENZ-15163407.HTML?PRINTPAGEDARTICLE=TRUE#PAGEINDEX_0. Zugegriffen: 28.07.2018

BaFin – Bundesanstalt für Finanzdienstleistungsaufsicht (2018). *Big Data trifft auf künstliche Intelligenz Herausforderungen und Implikationen für Aufsicht und Regulierung von Finanzdienstleistungen*. Bonn: BaFin.

Bünau, Paul. v. (2018). *Algorithmen und Transparenz*. https://legal-revolution.com/de/the-legal-revolutionary/itk/algorithmen-und-transparenz. Zugegriffen: 20.07.2018.

Bundesregierung (2016). *Positionspapier der Bundesrepublik Deutschland zum Regelungsumfeld für Plattformen, Online-Vermittler, Daten, Cloud Computing und die partizipative Wirtschaft (Konsultation der EU)*. https://www.bundesregierung.de/Content/DE/_Anlagen/BKM/2016/2016-04-22-positionspapier-plattformregulierung.pdf?__blob=publicationFile&v=2. Zugegriffen: 28.07.2018.

Busch, C. (2018). *Algorithmic Accountability*. Osnabrück: Universität Osnabrück.

Deutscher Ethikrat (2017). *Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung*. Berlin: Deutscher Ethikrat.

Döbel, I., Leis, M., Vogelsang, M. M., Neustroev, D., Petzka, H., Riemer, A., Rüping, S., Voss, A., Wegele, M., Welz, J. (2018). *Maschinelles Lernen*. München: Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V.

Geisberger, E., & Broy, M. (Hrsg.). (2012). *agendaCPS – Integrierte Forschungsagenda Cyber-Physical Systems*. München: acatech STUDIE.

Gräf, E., Lahmann, H., & Otto, P. (2018). *Die Stärkung der digitalen Souveränität, Deutsches Institut für Vertrauen und Sicherheit im Internet* – DIVSI (Hrsg.). http://www.iRights-Lab.de. Zugegriffen: 28.07.2018.

Kar, R. M., Thapa, B., & Parycek, P. (2004). *(Un)berechenbar – Algorithmen und Automatisierung in Staat und Gesellschaft*. Kompetenzzentrum Öffentliche IT (ÖFIT); Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS. Berlin.

Krüger, J. & Lischka, K. (2018). *Damit Maschinen den Menschen dienen – Lösungsansätze, um algorithmische Prozesse in den Dienst der Gesellschaft zu stellen*. Gütersloh: Bertelsmann Stiftung.

Mücklich, F. (2015). *Funktionswerkstoffe I & II*. Vorlesungsmanuskript Universität des Saarlands.

Otto, B. (2016). Digitale Souveränität: *Beitrag des Industrial Data Space*. München: Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e. V.

Rohde, N. (2018). *Gütekriterien für algorithmische Prozesse – Eine Stärken- und Schwächenanalyse ausgewählter Forderungskataloge*. Gütersloh: Bertelsmann Stiftung.

Schonschek, O., & Litzel, N. (2018). *Mehr Transparenz bei Künstlicher Intelligenz*. https://www.bigdata-insider.de/mehr-transparenz-bei-kuenstlicher-intelligenz-a-690269/. Zugegriffen: 28.07.2018.

Stockley, M. (2018). *Die Krux mit dem Machine Learning*. https://www.dotnetpro.de/diverses/krux-machine-learning-1482393.html. Zugegriffen: 28.07.2018.

Völz, H. (1999). *Das Mensch-Technik-System: physiologische, physikalische und technische Grundlagen;* Software und Hardware. Renningen-Malmsheim: Expertverlag.

Wachter, S. (2018). *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*. In Harvard Journal of Law & Technology.

Vöcking, B., Alt, H., Dietzfelbinger, M., Reischuk, R., Scheideler, C., Vollmer, H., & Wagner, D. (2008). *Taschenbuch der Algorithmen*. Berlin: Springer Verlag.

Zweig, K. A. (2018). *Überprüfbarkeit von Algorithmen*. https://algorithmwatch.org/de/zweites-arbeitspapier-ueberpruefbarkeit-algorithmen/. Zugegriffen: 20.07.2018.

[22] Busch 2018, S. 59

Zu diesem Thema könnten Sie auch folgende weitere Umsetzungshilfen interessieren:

- 1.1.4 Ethische Werte für die intelligente Software (inkl. KI)
- 1.1.7 Informationsblatt smartes Produkt
- 1.3.4 Autonome Softwaresysteme und Unternehmerverantwortung
- 1.3.5 Hersteller- und Unternehmerverantwortung in 4.0-Prozessen
- 2.1.2 Integration von intelligenter Software (inkl. KI) in die Organisation
- 2.3.1 Datensicherheit in 4.0-Prozessen
- 2.3.2 Datenschutz in 4.0-Prozessen
- 2.3.3 Datenqualität in 4.0-Prozessen
- 2.3.4 Betriebsvereinbarungen und Dienstvereinbarungen zu 4.0-Prozessen
- 2.5.3 Plattformökonomie

Herausgeber: „Offensive Mittelstand – Gut für Deutschland“ – Stiftung „Mittelstand – Gesellschaft – Verantwortung“
Kurfürsten-Anlage 62, 69115 Heidelberg, E-Mail: info@offensive-mittelstand.de; Heidelberg 2019