Digitale Souveränität: Soziotechnische Bewertung und Gestaltung von Anwendungen algorithmischer Systeme

Ernst A. Hartmann^(✉)

Institut für Innovation und Technik in der VDI/VDE Innovation + Technik
GmbH, Steinplatz 1, 10623 Berlin, Deutschland
hartmann@iit-berlin.de

Zusammenfassung. Digitale Souveränität ist eng verbunden mit der Handlungsfähigkeit von Individuen und Organisationen. Auf der Ebene des Individuums können handlungs- und kontrolltheoretische Konzepte zu einer genaueren Beschreibung und Operationalisierung digitaler Souveränität beitragen. Auf der Ebene von Arbeitssystemen stehen seit Jahrzehnten erprobte Verfahren zur Bewertung und Gestaltung zur Verfügung. Aufbauend auf diesen methodisch-theoretischen Grundlagen wird ein Konzept für die soziotechnische Bewertung von Anwendungen algorithmischer Systeme – insbesondere auf Künstlicher Intelligenz basierter Systeme – in der Arbeitswelt vorgestellt. Die Potenziale dieses Konzepts für die Auditierung und Zertifizierung solcher Anwendungen, in Kooperation mit den anwendenden Unternehmen, werden diskutiert.

Schlüsselwörter: Digitale Souveränität · Auditierung · Zertifizierung · Soziotechnische Bewertung · Organisationsgestaltung · Technikgestaltung · Arbeitsgestaltung

1 Handlungsfähigkeit und Kontrolle

Im vorliegenden Beitrag sollen Antworten auf die folgenden Fragen gesucht werden:

- Können algorithmische Systeme – darunter auch auf Künstlicher Intelligenz (KI) basierte Systeme – im Hinblick auf ihre Auswirkungen auf die digitale Souveränität in der Wirtschaft bzw. in der Arbeitswelt bewertet werden?
- Stehen für eine solche Bewertung angemessene wissenschaftliche und methodische Grundlagen zur Verfügung?
- Sollte sich eine solche Bewertung – auch im Sinne einer Auditierung oder Zertifizierung – auf die Produkte, also die algorithmischen Systeme selbst, beziehen und sich an deren Hersteller wenden, oder sollte sie die Anwendungen der Systeme im konkreten Nutzungskontext bei den Anwenderunternehmen betrachten?

Dieser Beitrag baut auf einem Kapitel des vorigen Themenbandes ‚Digitalisierung souverän gestalten‘ auf (Hartmann 2020). Dort wurden Grundlagen der Handlungs- und Kontrolltheorie dargestellt und hinsichtlich ihrer Eignung für die Ableitung

von Kriterien der Bewertung von KI-Systemen – oder allgemeiner: algorithmischen Systemen – im Hinblick auf die digitale Souveränität erörtert. Im hier vorliegenden Text soll diese Anwendung näher ausgearbeitet werden. Dazu werden im Folgenden die wesentlichen Gedankengänge des Vorläufertextes kurz zusammengefasst.

Der zentrale Gedankengang stützt sich dabei auf die Überlegung, dass sich digitale Souveränität ausdrückt in der *Handlungsfähigkeit* von Menschen oder Organisationen und der *Kontrolle*, die sie über ihre Umwelt und ihre Lebensbedingungen haben, speziell im Hinblick auf digitale Systeme, die diese Umwelt in Teilen konstituieren und mithilfe derer wiederum auch diese Kontrolle ausgeübt werden kann (Couture, Toupin 2019).

Aussagen zur menschlichen *Handlungsfähigkeit* und ihren Bedingungen macht die Handlungsregulationstheorie, auf die sich eine besonders im deutschsprachigen Raum sehr bedeutende Denkrichtung der Arbeitspsychologie stützt (Hacker 2005, 2010; Hacker, Richter 1990). Diese Theorie beschreibt Strukturen und Prozesse menschlicher Arbeitshandlungen und erlaubt es, Arbeitsbedingungen zu bewerten.

Weiterhin verweisen *Kontrolltheorien* (Karasek 1989) auf die Bedeutung der Kontrolle der Menschen über ihre Umwelt-, Lebens- und Arbeitsbedingungen als wesentliche Bedingung menschlicher Leistungsfähigkeit und Gesundheit; sie charakterisieren auch die objektiven Bedingungen, in denen eine solche Kontrolle entstehen kann.

Rainer Oesterreich (1981) verknüpfte nun Handlungs- und Kontrolltheorien mit seinen Konzepten des Handlungsraums sowie der *Effizienz* und der *Divergenz*. Der Handlungsraum ist eine abstrahierte Darstellung des Arbeitsumfeldes. Er besteht aus Situationen, in denen bestimmte Handlungen möglich sind, die dann wiederum – je nach gewählter Handlung – zu weiteren Situationen führen (ausführlichere Erläuterungen und Literaturverweise in Hartmann 2021).

Die *Effizienz* bezieht sich nun auf die Übergangswahrscheinlichkeiten, mit denen die in einer bestimmten Situation verfügbaren Handlungen zu definierten neuen Situationen führen. Idealerweise betragen alle diese Übergangswahrscheinlichkeiten 1,0; es führt also jede Handlung mit Sicherheit zu einer neuen Situation. Dieses Phänomen kann auch als Handlungssicherheit verstanden werden, also als ein Maß dafür, wie sicher sich ein arbeitender Mensch sein kann, mit einer bestimmten Handlung auch ein bestimmtes Ziel zu erreichen. Eingeschränkt wird die Effizienz, die Handlungssicherheit auf Seiten des Menschen etwa durch zu geringe Erfahrung, auf Seiten des technischen Systems durch mangelnde technische Zuverlässigkeit.

Ein zweiter Aspekt der Kontrolle, die Menschen über ihre Umwelt haben, ist die *Divergenz*. Sie hängt davon ab, wie viele neue Situationen durch Handlungen aus einer gegebenen Situation heraus erreicht werden können; die Divergenz ist eng verbunden mit den Konzepten des Handlungsspielraums und der Freiheitsgrade (Hacker 2005; Osterloh 1983). Sie wird begrenzt durch eine (zu) schmale Qualifikation des Menschen oder durch eine rigide, enge Benutzerführung durch das System.

Gerade im Hinblick auf komplexe, algorithmische Systeme ist die *Transparenz* des Systems eine ganz wesentliche Voraussetzung für Handlungsregulation und Kontrolle. Diese Transparenz kann unmittelbar gegeben sein durch inhärent transparente Systeme, oder sie muss indirekt hergestellt werden durch explizite Systemmodule, die dem Menschen eine Beschreibung und Erklärung der Prozesse und

Resultate des algorithmischen Systems geben können; solche Verfahren sind Gegenstand der Erklärbaren KI (Explainable Artificial Intelligence, XAI; vgl. Mueller et al. 2019).

Anhand dieser drei Dimensionen – Transparenz, Effizienz und Divergenz – soll im Folgenden ein Konzept soziotechnischer Souveränität entwickelt werden, auch als mögliche Grundlage der Bewertung und Gestaltung konkreter Anwendungen algorithmischer Systeme.

2 Soziodigitale Souveränität

Digitale Souveränität in der Wirtschaft bezieht sich, wie oben dargelegt, auf die Handlungsfähigkeit von Individuen und Organisationen und die Kontrolle, die sie über ihre Umgebungsbedingungen haben. Im Folgenden soll die Digitale Souveränität von Individuen – arbeitenden Menschen – im Vordergrund stehen, die für sie hier relevante Umwelt betrifft die Arbeitswelt.

Zur Charakterisierung von Arbeits(um)welten wurde das Konzept soziotechnischer Systeme entwickelt (Trist, Bamforth 1951; Ulich 2013). Nach dieser Konzeption sind drei Teilsysteme eines Arbeitssystems bzw. soziotechnischen Systems wesentlich:

- Der *Mensch* mit seinen Motiven und Bedürfnissen und seinen Kenntnissen und Fähigkeiten, die durch formales, non-formales und informelles Lernen erworben werden können.
- Die *Technik* in ihrem mehr oder weniger komplexen Aufbau aus mehr oder weniger heterogenen (mechanischen, elektronischen, optischen, fluidischen, hard- und softwarebasierten) Subsystemen, und zunehmend in ihrer komplexen Vernetzung verschiedenster, auch räumlich verteilter Systeme.
- Die *Organisation* mit den Aspekten der Aufgabenteilung und -kombination, der Aufbau- und Ablauforganisation, wobei formelle und informelle Organisationsstrukturen nebeneinander bestehen, wie auch der Organisations- bzw. Unternehmenskultur.

Kombiniert man diese drei soziotechnischen Teilsysteme mit den drei Kriterienbereichen digitaler Souveränität – Transparenz, Divergenz und Effizienz – ergibt sich die Neun-Felder-Matrix in Abb. 1.

Im Hinblick auf den *Menschen* hängt die *Transparenz* eines durch algorithmische Systeme geprägten Arbeitsumfeldes ab von IT-Grundkenntnissen, die es erlauben, die Arbeitsweise der Systeme zumindest in den Grundzügen verstehen zu können. Tiefe fachbezogene – darunter auch spezifische IT-bezogene – Kenntnisse und Fähigkeiten bestimmen das Maß an *Effizienz* – Handlungssicherheit –, mit der der Mensch agieren kann. Die *Divergenz* schließlich hängt ab von breiten, möglicherweise auch interdisziplinären Qualifikationen, die es erlauben, mehrere qualitativ unterschiedliche Wege zu bestimmten Zielen zu verfolgen.

Die *Organisation* – in allen ihren formellen und informellen Aspekten – kann durch Klarheit über Aufgaben, Rollen und Befugnisse zur *Transparenz* beitragen.

Neben anderen Faktoren wirkt das Maß an sozialer Unterstützung – selbst wiederum beeinflusst sowohl durch die Aufbauorganisation wie auch die Unternehmenskultur – auf die tatsächlich wirksame *Effizienz* der individuellen und kollektiven Handlungen. Ganz entscheidend für die *Divergenz* sind Handlungs- und Entscheidungsspielräume, die die Organisation den jeweiligen arbeitenden Menschen einräumt (Osterloh 1983).

Soziodigitale Souveränität
		Transparenz	Effizienz	Divergenz	
Was kann der Mensch?	Mensch	IT-Grundkenntnisse	Aufgabenbezogenes (digitales) Spezialwissen	Breites, interdisziplinäres (digitales) Spezialwissen	<ul style="list-style-type: none">• Formales ...• Non-formales ...• Informelles Lernen
Was fordert/erlaubt die Organisation?	Organisation	Transparenz über Aufgaben, Entscheidungsbefugnisse	Aufgabenzuschnitt, soziale Unterstützung	Entscheidungs-, Tätigkeits-, Handlungsspielräume	<ul style="list-style-type: none">• Formelle und ...• ... informelle Organisation• Unternehmenskultur
Was ermöglicht die Technik?	Technik	Transparenz, unmittelbare/mittelbare Erklärbarkeit	Technische Zuverlässigkeit, Robustheit, Resilienz	Eingriffsmöglichkeiten auf wählbaren Ebenen	<ul style="list-style-type: none">• Benutzeroberfläche• Offensichtliche und 'verborgene' Funktionen

Abb. 1. Merkmale soziodigitaler Souveränität

Die *Technik* schließlich kann sowohl durch ihren inneren Aufbau wie durch ihre Mensch-Technik-Schnittstelle mehr oder weniger zur *Transparenz* beitragen. Speziell für KI-basierte Systeme entstand eine umfangreiche Literatur zur XAI (vgl. Mueller et al. 2019). Der Beitrag der Technik zur *Effizienz* zeigt sich in einer hohen technischen Zuverlässigkeit, Robustheit und Resilienz. Die *Divergenz* beeinflusst Technik schließlich positiv, wenn sie etwa Eingriffsmöglichkeiten auf vom Menschen wählbaren Ebenen anbietet (vgl. Vicente, Rasmussen 1992; Lüdtke 2015).

Für die soziotechnische Bewertung von Arbeitssystemen stehen seit Jahrzehnten erprobte, detailliert ausgearbeitete Methoden und Kriterien zur Verfügung (z. B. Strohm, Ulich 1997; Ulich 2013; Jenderny et al. 2018).

Mit diesem Überblick sind die wesentlichen Konzepte zur Beschreibung soziotechnischer Souveränität dargestellt (für eine ausführlichere Diskussion vgl. Hartmann 2020). Im Folgenden sollen diese Konzepte in den Kontext aktueller Überlegungen zur Bewertung und Zertifizierung algorithmischer, auch KI-basierter Systeme eingeordnet werden. Dazu werden zunächst einige dieser aktuellen Vorschläge vorgestellt.

3 Konzepte zur Bewertung algorithmischer Systeme

In jüngerer Zeit wurden einige Vorschläge für Verfahren und Methoden zur Bewertung algorithmischer Systeme – insbesondere auch KI-gestützter Systeme – vorgebracht. Auf europäischer Ebene hat die High-Level Expert Group on Artificial Intelligence (AI HLEG) eine Bewertungsliste für vertrauenswürdige KI-Systeme – Assessment List for Trustworthy Artificial Intelligence (ALTAI) – vorgelegt (AI HLEG 2020). Diese Liste ist unterteilt in sieben Kriterienbereiche mit jeweiligen Unterkriteriengruppen:

1. Human Agency and Oversight
 - Human Agency and Autonomy
 - Human Oversight
2. Technical Robustness and Safety
 - Resilience to Attack and Security
 - General Safety
 - Accuracy
 - Reliability, Fall-back Plans and Reproducibility
3. Privacy and Data Governance
 - Privacy
 - Data Governance
4. Transparency
 - Traceability
 - Explainability
 - Communication
5. Diversity, Non-discrimination and Fairness
 - Avoidance of Unfair Bias
 - Accessibility and Universal Design
 - Stakeholder Participation
6. Societal and Environmental Well-being
 - Environmental Well-being
 - Impact on Work and Skills
 - Impact on Society at Large or Democracy
7. Accountability
 - Auditability
 - Risk Management

Abb. 2 zeigt diese Kriterien in Beziehung zu den Kriterien soziodigitaler Souveränität.

Abb. 2. Kriterienbereiche der Assessment List for Trustworthy Artificial Intelligence (ALTAI) in ihrer Beziehung zu den Merkmalen soziodigitaler Souveränität

Aus Abb. 2 wird deutlich, dass drei Kriterienbereiche der ALTAI sich sehr eng überlappen mit den drei Dimensionen soziodigitaler Souveränität:

- Die Kriterien ‚Transparenz‘ und ‚Transparency‘ sind in der Benennung identisch und auch inhaltlich sehr affin zueinander.
- Die Kriterien ‚Divergenz‘ – im Sinne der Möglichkeit der Wahl zwischen qualitativ verschiedenen Handlungsweisen als ein Kernmerkmal menschlicher Autonomie – einerseits und ‚Human Agency and Oversight‘ überlappen sich teilweise. Das erstgenannte Kriterium ist inhaltlich enger und fachwissenschaftlich wohl definiert (s. o.), das zweitgenannte deutlich breiter und dabei inhaltlich weniger konsistent unter Bezug auf spezifische wissenschaftliche Konzepte ausgearbeitet.
- Ähnlich verhält es sich bei den Kriterien ‚Effizienz‘ einerseits und ‚Technical Robustness and Safety‘ andererseits.

Ein genereller Unterschied besteht darin, dass die ALTAI nicht auf dem sozio-technischen Systemkonzept oder einer anderen systemischen Sichtweise beruht. Dies fällt besonders dort auf, wo an einzelnen Stellen auf nicht-technische Systemaspekte verwiesen wird – z. B. „Have the humans (...) been given specific training on how to exercise oversight?“ (AI HLEG 2020:8) –, ohne dass eine solche Betrachtung nach Teilsystemen systematisch über alle Kriterienbereiche hinweg durchgeführt würde.

Der Kriterienbereich ‚Privacy and Data Governance‘ fehlt bei den Kriterien der soziodigitalen Souveränität, weil er sich aus den entsprechenden wissenschaftlichen Grundlagen nicht (unmittelbar) ableiten lässt. In einer eventuellen zukünftigen Auditierung oder Zertifizierung sollten diese Aspekte als unabdingbare Voraussetzungen der digitalen Souveränität aber unbedingt integriert werden.

Die Kriterienbereiche ‚Diversity, Non-discrimination and Fairness‘ und ‚Societal and Environmental Well-being‘ der ALTAI beziehen sich teilweise auf Sachverhalte,

die für Anwendungen in der Arbeitswelt nicht immer von Bedeutung sind. So ist etwa die Nichtdiskriminierung durch Algorithmen insbesondere dort relevant, wo Personen durch algorithmische Entscheidungen in unfairer Weise betroffen sein könnten, wie etwa bei algorithmischen Entscheidungen zur Kreditvergabe. Solche Implikationen sind bei KI-Anwendungen in der Arbeitswelt nicht immer gegeben. Andererseits ist die Senkung von Nutzungshürden für diverse Nutzergruppen – z. B. auch für Personen mit spezifischen Bedarfen und Einschränkungen – auch für Systeme in Arbeitskontexten fast immer relevant; hierzu liegt auch eine substanzielle Forschungsliteratur vor (für einen Überblick vgl. z. B. Lim et al. 2021).

Accountability kann schließlich als ein ‚Meta-Kriterium‘ betrachtet werden, das die Auslegung eines zu bewertenden Systems (was immer das System im konkreten Fall sein mag) im Hinblick auf die Ermöglichung einer Bewertung, Auditierung, Zertifizierung fordert. Dieser Aspekt sollte bei allen Bewertungen, Auditierungen, Zertifizierungen beachtet werden.

Im Whitepaper zur KI-Zertifizierung des Fraunhofer-Instituts für Intelligente Analyse- und Informationssysteme IAIS (FhG-IAIS) in Kooperation mit der Universität Bonn (Institut für Philosophie) und der Universität zu Köln (Institut für Strafrecht und Strafprozessrecht) wird die Arbeit der AI HLEG aufgegriffen und es werden – auf der „Basis von Ethik und Recht“ – sehr ähnliche Kriterien vorgeschlagen (Cremers et al. 2019:15 ff.):

- „Autonomie & Kontrolle: Ist eine selbstbestimmte, effektive Nutzung der KI möglich?
- Fairness: Behandelt die KI alle Betroffenen fair?
- Transparenz: Sind Funktionsweise und Entscheidungen der KI nachvollziehbar?
- Verlässlichkeit: Funktioniert die KI zuverlässig und ist sie robust?
- Sicherheit: Ist die KI sicher gegenüber Angriffen, Unfällen und Fehlern?
- Datenschutz: Schützt die KI die Privatsphäre und sonstige sensible Informationen?“

Im Vergleich zu den oben beschriebenen Kriterien der ALTAI fehlen hier die Kriterien 6 ‚Societal and Environmental Wellbeing‘ und 7 ‚Accountability‘, dafür wird das Kriterium 2 ‚Technical Robustness and Safety‘ aufgeteilt in die Kriterien ‚Verlässlichkeit‘ und ‚Sicherheit‘. Mittlerweile liegt auch ein auf dem Whitepaper basierender ‚Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz‘ vor (Fraunhofer IAIS 2021).

In der ‚Deutschen Normungsroadmap Künstliche Intelligenz‘ werden zum Thema ‚Qualität, Konformitätsbewertung und Zertifizierung‘ (DIN/DKE 2020:79 ff.) wiederum die oben genannten Kriterien des Whitepapers des Konsortiums aus IAIS und den Universitäten Bonn und Köln (Cremers et al. 2019) zitiert.

Insgesamt fällt bei diesen Vorschlägen – insbesondere in Bezug auf die AI HLEG und das Konsortium um das FhG-IAIS – auf, dass einerseits die Bedeutung einer interdisziplinären Betrachtungsweise sehr betont wird, andererseits in beiden Fällen sich diese Interdisziplinarität im Wesentlichen auf das Dreieck Informatik, Jura, und Philosophie beschränkt. Weiterhin wird in beiden Fällen eine starke Fokussierung auf den Menschen einerseits und auf gesellschaftliche Belange andererseits gefordert – umso mehr verblüfft die (fast vollständige, s. u.) Abwesenheit

human- oder sozialwissenschaftlicher Kompetenz in den jeweiligen Konsortien bzw. Arbeitsgruppen. Das FhG-IAIS kooperiert mit jeweils einer juristischen und philosophischen Partner-Institution. Unter den 55 Mitgliedern der AI HLEG sind – neben Repräsentanten und Repräsentantinnen der Industrie, der Gewerkschaften und weiterer Interessensgruppen – von wissenschaftlicher Seite wiederum die Fächer Informatik, Jura und Philosophie (insbesondere Ethik) gut vertreten. Human- und sozialwissenschaftliche Kompetenz im engeren Sinne fehlt hingegen fast völlig, mit der (bedingten) Ausnahme von Sabine Theresia Köszegi, Professorin für Arbeitswissenschaft und Organisation am Institut für Managementwissenschaften der Technischen Universität Wien, die über einen betriebswirtschaftlichen Bildungshintergrund verfügt.

Aus dieser Diskussion lässt sich festhalten:

- In den drei für die Arbeitswelt wesentlichen Bereichen ‚Transparenz‘ (‚Transparency‘), ‚Effizienz‘ (‚Technical Robustness and Safety‘) und ‚Divergenz‘ (‚Human Agency and Oversight‘) gibt es wesentliche Überlappungen zwischen den arbeitswissenschaftlich begründeten soziodigitalen Kriterien und entsprechenden Kriterienbereichen der ALTAI.
- Dabei können die soziotechnisch begründeten Kriterien wesentlich dazu beitragen, die recht lückenhafte human- und sozialwissenschaftliche Begründung der ALTAI-Kriterien zu ergänzen.
- Umgekehrt liefern die Kriterien der AI HLEG und des FhG-IAIS-Konsortiums viele Hinweise für weitere Kriterien, die aus einer rein soziotechnischen Betrachtung (noch) nicht ableitbar, aber für eine Bewertung von Arbeitssystemen unerlässlich sind (wie etwa Datenschutz/Privacy).
- Einige Kriterienbereiche der ALTAI beziehen sich auf breitere gesellschaftliche Fragestellungen, die für arbeitsweltbezogene Betrachtungen nicht immer relevant sind und daher fallweise hinzugezogen werden sollten.

4 Bewertungsgegenstände: KI-Systeme oder deren Anwendungen im Nutzungskontext?

Konformitätsprüfung ist der Oberbegriff für Prozesse des Evaluierens (Bewertens) von Gegenständen. Solche Evaluationen können zu einer Zertifizierung führen, wenn das Ergebnis der Konformitätsprüfung von einer dritten Seite (Third Party) bestätigt wird, die unabhängig ist vom Anbieter (erste Seite/First Party) des in Frage stehenden Gegenstandes und kein Interesse an der Nutzung (Nutzer als zweite Seite/Second Party) des Gegenstandes hat (DIN/DKE 2020:83).

Die Deutsche Normungsroadmap KI (DIN/DKE 2020) ebenso wie das FhG-IAIS-Whitepaper bzw. der darauf basierende Leitfaden (Cremers et al. 2019; Fraunhofer IAIS 2021) zielen auf Zertifizierung durch unabhängige Dritte. Das ALTAI-Assessmenttool ist demgegenüber für eine Selbsteinschätzung durch Hersteller/Anbieter (First Party) oder Nutzer (Second Party) ausgelegt.

Als Gegenstände der Konformitätsprüfung kommen grundsätzlich in Frage (DIN/DKE 2020:81):

1. „Produkt (z. B. Hard-/Software)
2. Prozess
3. System
4. Dienstleistung
5. Managementsystem
6. Person
7. Information (z. B. Deklarationen, Behauptungen, Vorhersagen)“

Auch Kombinationen dieser Gegenstände sind möglich.

Eine weitere Fragestellung bezieht sich darauf, inwieweit sich die Bewertung bzw. Zertifizierung auf das KI-gestützte Produkt selbst beziehen soll – dann wäre der Hersteller Kunde des Verfahrens – oder auf die Nutzung des KI-basierten Systems in einem Anwendungskontext¹ als Teil eines soziotechnischen Systems – mit dem Anwenderbetrieb als Kunden.

Die ALTAI ist hier eher offen und soll verstanden werden als „Assessment List to help assess whether the AI system that is being developed, deployed, procured or used, adheres to the seven requirements of Trustworthy Artificial Intelligence“ (AI HLEG 2020:3).

Die Überlegungen der KI-Normungsroadmap und des FhG-IAIS-Whitepapers bzw. -Leitfadens zielen (eher) auf Zertifizierungen der KI-Produkte selbst.

Axel Mangelsdorf, Peter Gabriel und Martin Weimer (2021) diskutieren kritisch die Vor- und Nachteile solcher KI-Produktzertifizierungen. Als Pro-Argumente werden u. a. genannt:

- Solche Zertifizierungen könnten zu mehr Rechtssicherheit für Entwickler und Anwender führen.
- Weiterhin könnten sie Wettbewerbsvorteile für insbesondere europäische KI-Anbieter hervorbringen.

Als Contra-Argumente werden dem u. a. die folgenden gegenübergestellt

- Die Kriterien seien kaum oder gar nicht prüfbar – zumindest nicht ohne Bezug auf spezifische Anwendungskontexte.
- Die Zertifizierungskosten seien insbesondere für kleine und mittlere Unternehmen als KI-Anbieter zu hoch.

Eine finale Bewertung nehmen die Autoren nicht vor, weil dies zum jetzigen Zeitpunkt noch nicht möglich sei.

¹ Zu einer begrifflichen Verwirrung kann hier beitragen, dass auch KI-Systeme selbst gelegentlich, in Analogie zum Begriff der Anwendungssoftware, als ‚KI-Anwendungen‘ bezeichnet werden. Im hier vorliegenden Text ist mit ‚Anwendung‘ immer die Nutzung eines solchen Systems in einem konkreten Anwendungskontext als Teilsystem eines soziotechnischen Systems – das auch organisationale Strukturen und die dort arbeitenden Menschen umfasst – gemeint.

Unabhängig von den Vorzügen und Nachteilen einer produktbezogenen Zertifizierung sollen im Folgenden einige Implikationen einer anwendungsbezogenen, auf konkrete Nutzungskontexte im Rahmen soziotechnischer Systeme bei den Anwenderunternehmen ausgerichteten Bewertung diskutiert werden.

Eine erste Überlegung bezieht sich darauf, wann ein Produkt oder System eigentlich ‚fertig‘ oder ‚funktional wirksam entstanden‘ ist. Viele Produkte entstehen in ihrer (vorläufig) finalen Form erst beim Kunden bzw. Anwender, durch kundenspezifische Variantenkonstruktion, Konfiguration oder Parametrisierung. Insbesondere KI-basierte Systeme werden mitunter beim/vom Kunden trainiert.

Auch jenseits der Frage nach der betriebsbereiten Fertigstellung des unmittelbaren technischen Systems selbst kann – im soziotechnischen Kontext – nach den Bedingungen gefragt werden, unter denen das technische System funktional wirksam wird. So kann etwa ein technisches System durch hohe Freiheitsgrade der Mensch-Technik-Interaktion grundsätzlich diversitätsförderlich ausgelegt sein. Dies wird aber möglicherweise funktional nicht wirksam, wenn organisationale Regelungen dem Menschen nur eine bestimmte Vorgehensweise erlauben, oder diese Person von ihrer Qualifikation her nur eine Methode beherrscht.

Aus dieser Perspektive entstehen technische Systeme funktional wirksam erst im Nutzungskontext, als Teilsysteme soziotechnischer Systeme. Dieser Umstand würde bei einer soziotechnischen Bewertung, eventuell auch Zertifizierung, des Systems im Anwendungskontext vollständig berücksichtigt werden können.

Die Kunden einer solchen Anwendungsbewertung wären somit keine KI-Anbieter wie bei einer Produktzertifizierung, sondern Unternehmen unterschiedlichster Branchen, die algorithmische oder KI-basierte Systeme einsetzen möchten. Besonders interessant wäre eine solche soziotechnische Bewertung bzw. Zertifizierung für solche Unternehmen, die an einer sozialpartnerschaftlichen Einführung und einem kooperativen Betrieb dieser Systeme interessiert sind und dafür objektivierbare Grundlagen suchen.

Für die konkrete Ausgestaltung eines solchen Verfahrens könnte ein kooperativer, partizipativer Ansatz mit geteilter Verantwortung gewählt werden. Dabei würde die Analyse des soziotechnischen Systems von innerbetrieblichen Akteuren gemeinsam mit externen Experten durchgeführt. Für die abschließende Bewertung könnte allerdings die externe Institution die volle Verantwortung übernehmen.

5 Fazit und Ausblick

Eingangs wurden folgende Fragen aufgeworfen:

- Können algorithmische Systeme – darunter auch KI-basierte Systeme – im Hinblick auf ihre Auswirkungen auf die digitale Souveränität in der Wirtschaft bzw. in der Arbeitswelt bewertet werden?
- Stehen für eine solche Bewertung angemessene wissenschaftliche und methodische Grundlagen zur Verfügung?
- Sollte sich eine solche Bewertung – auch im Sinne einer Auditierung oder Zertifizierung – auf die Produkte, also die algorithmischen Systeme, beziehen und sich an deren Hersteller wenden, oder sollte sie die Anwendungen der Systeme im konkreten Nutzungskontext bei den Anwenderunternehmen betrachten?

Diese Fragen können nun wie folgt beantwortet werden:

- Algorithmische bzw. KI-basierte Systeme können im Hinblick auf ihre Auswirkungen auf die digitale Souveränität in der Wirtschaft bzw. in der Arbeitswelt bewertet werden; es liegen auch schon diverse Vorschläge für Bewertungsverfahren vor, die ebenso auf Arbeitssysteme anwendbar sind.
- Psychologische Handlungs- und Kontrolltheorien und soziotechnische Bewertungs- und Gestaltungsmodelle stehen als theoretisch gut begründete und empirisch bewährte Grundlagen zur Verfügung. Sie können auch für die bislang noch in weiten Teilen fehlende human- und sozialwissenschaftliche Begründung der aktuell diskutierten Bewertungsverfahren herangezogen werden.
- Die Möglichkeit und der Nutzen produktbezogener Bewertungs- und Zertifizierungsverfahren wird kontrovers diskutiert. Unabhängig davon lassen sich gute Gründe für ein anwendungsbezogenes Herangehen finden: Viele Aspekte der tatsächlichen Wirkung und Wirksamkeit von Gestaltungsvarianten technischer Systeme lassen sich erst im soziotechnischen Nutzungskontext (gut) bewerten; weiterhin können solche Verfahren viele Anwenderunternehmen bei der sozialpartnerschaftlichen Einführung algorithmischer bzw. KI-basierter Systeme auf einer objektivierbaren Basis unterstützen.

Als nächste Schritte sollten

- Prototypen soziotechnischer Bewertungsverfahren für algorithmische bzw. KI-basierte Systeme im Nutzungskontext auf der Basis vorhandener theoretischer und methodischer Grundlagen, wie in diesem Beitrag umrissen, ausgearbeitet und mit operationalisierten, mess- bzw. beobachtbaren Kriterien unterlegt,
- Geschäftsmodelle für solche Verfahren entwickelt und
- Möglichkeiten und Nutzenpotenziale eines weiteren Ausbaus bis hin zu Zertifizierungsverfahren eruiert werden.
- Im Zuge dieser weiteren Entwicklung müssen die Verfahren iterativ in realen Anwendungsfällen getestet und weiterentwickelt werden.

Literatur

- AI HLEG: High-level expert group on artificial intelligence: the assessment list for trustworthy artificial intelligence (ALTAI). <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (2020). Zugriffen: 21. Juli 2021.
- Couture, S., Toupin, S.: What does the notion of ‘sovereignty’ mean when referring to the digital? *New Media Soc.* **21**(10), 2305–2322 (2019)
- Cremers, A.B., Englander, A., Gabriel, M., Hecker, D., Mock M., Poretschkin, M., Julia Rosenzweig, J., Rostalski, F., Volmer, J., Voosholz, J., Voss, A., Wrobel, S.: Vertrauenswürdiger Einsatz von Künstlicher Intelligenz – Handlungsfelder aus philosophischer, ethischer, rechtlicher und technologischer Sicht als Grundlage für eine Zertifizierung von Künstlicher Intelligenz. Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS. https://www.iais.fraunhofer.de/content/dam/iais/KINRW/Whitepaper_KI-Zertifizierung.pdf (2019). Zugriffen: 7. Aug. 2021

- DIN/DKE: Deutsches Institut für Normung, Deutsche Kommission für Elektrotechnik. Deutsche Normungsroadmap Künstliche Intelligenz, Berlin. <https://www.din.de/resource/blob/772438/6b5ac6680543eff9fe372603514be3e6/normungsroadmap-ki-data.pdf> (2020). Zugriffen: 7. Aug. 2021
- Fraunhofer IAIS (Hrsg.): Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz. https://www.iais.fraunhofer.de/content/dam/iais/fb/Kuenstliche_intelligenz/ki-pruefkatlog/202107_KI-Pruefkatalog.pdf (2021). Zugriffen: 8. Aug. 2021
- Hacker, W.: Software-Ergonomie; Gestalten Rechnergestützter Geistiger Arbeit?! In: Schönpflug, W., Wittstock, M. (Hrsg.) Software-Ergonomie '87 Nützen Informationssysteme dem Benutzer? Berichte des German Chapter of the ACM. Vieweg+Teubner, Wiesbaden (1987)
- Hacker, W.: Allgemeine Arbeitspsychologie, 2., vollständig überarbeitete und ergänzte Aufl. Huber, Bern (2005)
- Hacker, W.: Psychische Regulation von Arbeitstätigkeiten. In: Kleinbeck, U., Schmidt, K.-H. (Hrsg.) Arbeitspsychologie. Hogrefe, Göttingen (2010)
- Hacker, W., Richter, P.: Psychische Regulation von Arbeitstätigkeiten – ein Konzept in Entwicklung. In: Frei, F., Udris, I. (Hrsg.) Das Bild der Arbeit. Huber, Bern (1990)
- Hartmann, E.A.: Digitale Souveränität in der Wirtschaft – Gegenstandsbereiche, Konzepte und Merkmale. In: Hartmann, E.A. (Hrsg.) Digitalisierung souverän gestalten. Springer Vieweg, Wiesbaden (2020)
- Jenderny, S., Foullois, M., Kato-Beiderwieden, A.-L., Bansmann, M., Wöste, L., Lamß, J., Maier, G. W., Röcker, C.: Development of an instrument for the assessment of scenarios of work 4.0 based on socio-technical criteria. In: PETRA '18: Proceedings of the 11th Pervasive Technologies Related to Assistive Environments. Conference June 2018 (2018)
- Karasek, R.: Control in the workplace and its health-related aspects. In: Sauter, S.L., Hurrell, J.J., Cooper, C.L. (Hrsg.) Job Control and Worker Health, S. 129–160. Wiley, Chichester (1989)
- Lim, Y., Giacomini, J., Nickpour, F.: What is psychosocially inclusive design? A definition with constructs. *Des. J.* **24**(1), 5–28 (2021)
- Lüdtke, A.: Wege aus der Ironie in Richtung ernsthafter Automatisierung. In: Botthof, A., Hartmann, E.A. (Hrsg.) Zukunft der Arbeit in Industrie 4.0. Springer Vieweg, Heidelberg (2015)
- Mangelsdorf, A., Gabriel, P., Weimer, M.: Die Zertifizierung von KI: Mehr Sicherheit für alle – oder unnötiger Ballast? iit-perspektive. <https://www.iit-berlin.de/publikation/die-zertifizierung-von-ki-mehr-sicherheit-fuer-alle-oder-unnoetiger-ballast/> (2021). Zugriffen: 29. Juli 2021
- Mueller, S.T., Hoffman R.R., Clancey, W., Emrey, A., Klein, G.: Explanation in human-AI systems: a literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI. <https://arxiv.org/ftp/arxiv/papers/1902/1902.01876.pdf> (2019)
- Oesterreich, R.: Handlungsregulation und Kontrolle. Urban & Schwarzenberg, München (1981)
- Oesterreich, R.: Zur Analyse von Planungs- und Denkprozessen in der industriellen Produktion – Das Arbeitsanalyseinstrument VERA. *Diagnostica* **30**(3), 216–234 (1984)
- Osterloh, M.: Handlungsspielräume und Informationsverarbeitung. Huber, Bern (1983)
- Strohm, O., Ulich, E.: Unternehmen arbeitspsychologisch bewerten. Ein Mehr-Ebenen-Ansatz unter besonderer Berücksichtigung von Mensch, Technik, Organisation. vdf, Zürich (1997)
- Stubbe, J.: Von digitaler zu soziodigitaler Souveränität. In: Wittpahl, V. (Hrsg.) Digitale Souveränität. Bürger, Unternehmen, Staat. Springer Vieweg Open, Berlin (2017)
- Trist, E.L., Bamforth, K.W.: Some social and psychological consequences of the longwall method of coal-getting: an examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Hum. Relat.* **4**, 3–38 (1951)

- Ulich, E.: Arbeitssysteme als soziotechnische Systeme – Eine Erinnerung. J. Psychol. Alltagshandelns **6**(1), 4–12 (2013)
- Vicente, K.J., Rasmussen, J.: Ecological interface design: theoretical foundations. IEEE Trans. Syst. Man Cybern. **22**(4), 589–606 (1992)

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.