INSTITUT FÜR INNOVATION UND TECHNIK
Hochschule Aalen
nexus
DENK FABRIK DIGITALE ARBEITS-GESELLSCHAFT
Gefördert durch:
Bundesministerium für Arbeit und Soziales
aufgrund eines Beschlusses des Deutschen Bundestages

Ein KI-Cockpit für Beschäftigte

Künstliche Intelligenz am Arbeitsplatz verstehen, souverän einsetzen und kontrollieren

Autor:innen
Aschenbrenner, Bottel, Colloseus, Guagnin, Hubel, Jantzen, Kempen, Peters, Saleh, Sell

KI cockpit

AUTOR:INNEN

Prof. Dr. Doris Aschenbrenner, Hochschule Aalen
Matthias Bottel, nexus – Institut für Kooperationsmanagement u. interdisziplinäre Forschung
Cecilia Colloseus, Hochschule Aalen
Dr. Daniel Guagnin, nexus – Institut für Kooperationsmanagement u. interdisziplinäre Forschung
Nikolas Hubel, Institut für Innovation und Technik
Lisa Jantzen, Hochschule Aalen
Prof. Dr. Regina Kempen, Hochschule Aalen
Dr. Robert Peters, Institut für Innovation und Technik
Faten Saleh, Institut für Innovation und Technik
Andrea Sell, Hochschule Aalen

Das Autor:innenteam bedankt sich bei den Mitgliedern des Projektbeirats, namentlich bei Frau Lajla Fetic und Herrn Dietmar Kuttner, für die konstruktive Kommentierung.

GESTALTUNG

Martijn Verbeij
StudioConvex.nl
Design & Event Solutions

BILDNACHWEIS (TITEL)

Martijn Verbeij
StudioConvex.nl
Design & Event Solutions

IMPRESSUM

Institut für Innovation und Technik (iit)
in der VDI/VDE Innovation + Technik GmbH
Steinplatz 1
10623 Berlin

www.iit-berlin.de

Kontakt
Faten Saleh
030 310078-241
saleh@iit-berlin.de

Berlin, November 2024

INHALTSVERZEICHNIS

ABSTRACT

Künstliche Intelligenz (KI) birgt großes Potenzial, die Arbeitsqualität zu verbessern, die Produktivität zu steigern und Mitarbeitende zu entlasten. Damit dieses Potenzial voll ausgeschöpft werden kann, ist ein menschenzentrierter Ansatz bei der Entwicklung und dem Einsatz von KI-Technologien erforderlich. Die KI-Verordnung der EU legt hierfür klare Anforderungen fest, insbesondere für Hochrisiko-KI-Systeme, bei denen menschliche Aufsicht und Kontrolle durch qualifizierte Operator:innen sichergestellt sein muss.

Dieses Whitepaper präsentiert zentrale Erkenntnisse aus dem Projekt „KI-Cockpit“, das die exemplarische Umsetzung der europäischen KI-Verordnung in der betrieblichen Praxis untersucht.[1] Das Projekt demonstriert, wie KI-Systeme so gestaltet werden können, dass menschliche Transparenz, Kontrolle und Steuerung gewährleistet sind. Der „Human in Command“-Ansatz, der im Mittelpunkt des Projekts steht, integriert den Menschen aktiv in die Überwachung und Gestaltung von KI-Systemen, wodurch Risiken minimiert und die Akzeptanz von KI gestärkt werden.

Mit dem KI-Cockpit wird eine technische Lösung entwickelt, die Transparenz, Fairness und Kontrollmöglichkeiten durch Performance-Monitoring, Testfälle und kontextspezifische Autonomiestufen bietet. Diese Funktionen stellen sicher, dass technologische Effizienzpotenziale genutzt werden, ohne dass Mitarbeitende Kontrolle und Verantwortung abgeben müssen.

Zudem zeigt das Whitepaper auf, wie durch Standardisierungsprozesse und politische Maßnahmen, etwa den AI Pact[2] und Förderprogramme, die Einführung menschenzentrierter KI-Anwendungen gestärkt werden kann. Dies fördert nicht nur den ethischen und verantwortungsvollen Umgang mit KI-Technologien, sondern trägt auch dazu bei, die Chancen hybrider Intelligenz in verschiedenen Branchen zu realisieren.

[1] Das Projekt wird von nexus, der Hochschule Aalen, dem Institut für Arbeitswissenschaft und Technologiemanagement (IAT) der Universität Stuttgart, Chemistree GmbH, Starwit Technologies GmbH sowie der Caritas Dortmund durchgeführt und vom Bundesministerium für Arbeit und Soziales (BMAS) gefördert.

[2] https://digital-strategy.ec.europa.eu/en/policies/ai-pact.

1. EINLEITUNG

Große Versprechen gehen mit dem aktuellen Hype um künstliche Intelligenz (KI) einher: KI soll Beschäftigte entlasten, die Arbeitsqualität verbessern, die Produktivität erhöhen und den demografischen Herausforderungen entgegenwirken – kurz gesagt: einen Mehrwert für die Gesellschaft schaffen. Dies ist auch das Ziel der ressortübergreifenden Initiative *Civic Coding* – KI für das Gemeinwohl.[3] Gleichzeitig gibt es Bedenken bezüglich einer ausreichenden Kontrolle über KI-Systeme, die sich in der lebendigen Debatte rund um die KI-Verordnung der EU widerspiegeln. Eine entscheidende Voraussetzung für die ausreichende Kontrolle ist, dass Entwicklung, Gestaltung und Einsatz der Technologie menschenzentriert sowie eng an den Bedürfnissen und Fähigkeiten der Beschäftigten orientiert sind. Dazu gehört, dass Beschäftigte KI-Systeme im Arbeitsalltag verstehen, überwachen und steuern können. KI ist nur dann eine sinnvolle Unterstützung, wenn die Entscheidung über ihren Einsatz beim Menschen liegt und es eine Möglichkeit gibt, die Ergebnisse von algorithmischen Systemen zu kontrollieren. Wie die Zusammenarbeit von Menschen mit KI-Systemen in der Praxis aussieht, welche Vorteile KI für Wertschöpfung und Beschäftigung schafft, hängt wesentlich davon ab, ob es uns gelingt, die Technologie so zu gestalten, dass sie sich dem Menschen anpasst und nicht umgekehrt. Auf internationaler Ebene ist dieser Aspekt bereits angekommen und wird in der Literatur auch als „hybride Intelligenz“ bezeichnet.[4]

Das Forschungsprojekt KI-Cockpit möchte daher die Frage beantworten, wie die transparente Gestaltung von KI-Technologien und die Möglichkeit, sie für Menschen kontrollierbar zu machen, in der betrieblichen Praxis gelingen können. Ziel ist es, Beschäftigten einen Überblick über die Funktionsweise und die Entscheidungen des Systems zu vermitteln und sie zu befähigen, gut informiert und selbstständig über die Notwendigkeit von Eingriffen in maschinelle Abläufe zu entscheiden. Damit sollen insbesondere gesellschaftliche Risiken von KI-Technologien wie Verzerrungen und Diskriminierungen minimiert und die Akzeptanz ihrer Nutzung in der Arbeitswelt gesteigert werden. Das Projekt leistet einen Beitrag, damit der Anspruch eines Paradigmenwechsels von technologie- zu menschenzentrierter Technologiegestaltung[5] Wirklichkeit wird. Erreicht werden soll dies mit der im Projekt entwickelten Open-Source-Software „KI-Cockpit“.

KI-Cockpit als Antwort auf KI-Regulierung – Gesellschaftlicher und ökonomischer Nutzen

Aus einem breiten gesellschaftlichen Diskurs heraus, an dem Akteure aus Zivilgesellschaft, Wissenschaft, Wirtschaft und Politik seit vielen Jahren teilhaben, ist in Europa der weltweit erste, umfassende Regulierungsrahmen für KI entstanden. Die KI-Verordnung der EU legt Rahmenbedingungen für das Inverkehrbringen und den Einsatz von KI fest und betont dabei die Bedeutung menschlicher Letztentscheidung. Die gesetzlichen Regelungen der EU-KI-Verordnung sollen mit dem Projekt in konkrete Beispiele für eine erfolgreiche Umsetzung übersetzt werden. Menschen, die mit KI-Systemen arbeiten, soll das KI-Cockpit einen verantwortungsvollen Einsatz dieser Technologie ermöglichen. Dazu werden im Forschungsprojekt für drei „Hochrisiko“-Anwendungsbereiche – Personalwesen, smarte Kommune und smarte Pflege – eine Software-Implementierung sowie ein Vorgehensmodell entwickelt, das Organisationen bzw. Unternehmen bei der mit der KI-Verordnung konformen Entwicklung und Einführung von KI-Systemen unterstützt. Mit diesem Ansatz soll aus hohen rechtlichen und ethischen Ansprüchen (sozio-)technische Wirklichkeit werden. Gleichzeitig wird durch die besser ausgestaltete Mensch-Technik-Interaktion der Einsatz von KI attraktiver, Arbeitsprozesse werden effizienter und hybride Intelligenz kann so die Wettbewerbsfähigkeit von Unternehmen langfristig stärken.

[3] Vgl. https://www.civic-coding.de/angebote/publikationen/forschungsbericht.

[4] Akata et al., „A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence“, in Computer, vol. 53, no. 8, pp. 18-28, Aug. 2020, doi: 10.1109/MC.2020.2996587.

[5] https://www.denkfabrik-bmas.de/fileadmin/Downloads/Publikationen/Arbeiten_mit_Kuenstlicher_Intelligenz_bf.pdf

2. DAS „KI-COCKPIT“-FORSCHUNGSPROJEKT FÜR MENSCHENZENTRIERTE KI

Das Projekt *„KI-Cockpit – Implementierung von Praxisbeispielen für ‚Human in Command‘“* erprobt einen neuen Ansatz zur Beaufsichtigung und Kontrolle von KI-Systemen und stellt dabei den Menschen in den Mittelpunkt. Es liefert eine unmittelbare Antwort auf die gegenwärtig stark diskutierte Frage, wie die Anforderungen der EU-KI-Verordnung für Betreiber von KI-Systemen im Sinne der Zielstellung der europäischen KI-Regulierung umgesetzt werden können. Während sich die wissenschaftliche und politische Debatte hier gegenwärtig vor allem auf Fragen von Governance und juristische Aspekte fokussiert, zielt das Projekt auf eine praktische Lösung an der Stelle, auf die es vor allem ankommt: die Mensch-Maschine-Schnittstelle. Damit ist der Teil von KI-Systemen gemeint, mit dem wir als menschliche Nutzer:innen unmittelbar in Berührung kommen. Diese Schnittstelle ist essentiell für eine effektive Zusammenarbeit zwischen Mensch und künstlicher Intelligenz.

Das Forschungsprojekt adressiert diese Herausforderungen mithilfe eines interdisziplinären Ansatzes. Dazu wird auf technischer Seite ein *KI-Cockpit* als zentrale Überwachungs- und Steuerungssoftware für die menschliche Aufsicht sowie ein *Transparenz-Interface* für die Befähigung zu informierten Entscheidungen als Open-Source-Anwendung entwickelt. Während das KI-Cockpit unmittelbar an die Forderungen der KI-Verordnung nach Implementierung geeigneter Mensch-Maschine-Schnittstellen anknüpft, geht das Projekt mit dem *Transparenz-Interface* darüber hinaus: Die Informationspflichten der Anbieter und Betreiber beschränken sich gegenüber den Endnutzer:innen aus der KI-Verordnung (Art. 50 [1]) auf ein Minimalmaß, nämlich darüber zu informieren, dass mit einem KI-System interagiert wird. Zusätzlich ist es aber essenziell, Endnutzer:innen eine informierte Entscheidung für oder gegen die Nutzung des Systems zu ermöglichen.

Als Orientierung für die Ausgestaltung der im Projekt entwickelten Lösungen dienen die besonders hohen Anforderungen für Hochrisiko-KI-Systeme, beispielsweise im Personalbereich. Auf dieser Basis werden Lösungen entwickelt, die auch in anderen Anwendungskontexten anschlussfähig sind. Das KI-Cockpit kann so als Blaupause für ein weites Feld von KI-Anwendungen dienen und als Katalysator für die menschenzentrierte Gestaltung von KI-Entwicklung in Deutschland und darüber hinaus wirken.

2.1 Vier verschiedene Nutzer:innengruppen der menschlichen Aufsicht für KI-Systeme

Kurzgefasst

- Artikel 14 der EU-KI-Verordnung fordert menschliche Aufsicht für den Betrieb von Hochrisiko-KI-Systemen, um Risiken für Gesundheit, Sicherheit und Grundrechte zu minimieren.
- Softwarehersteller müssen gemäß der KI-Verordnung Mensch-Maschine-Schnittstellen umsetzen; die operative Aufsicht wird von qualifizierten KIC-Operator:innen durchgeführt.
- Das KI-Cockpit ermöglicht, diese Aufsicht soziotechnisch zu realisieren, indem es KIC-Operator:innen befähigt, KI-Entscheidungen zu verstehen, zu kontrollieren und zu beeinflussen.
- Das Projekt unterscheidet drei Rollen: KIC-Operator:in (Expert:in für die Bedienung des KI-Cockpits), Endnutzer:in (Anwender:in ohne Zugriff auf das Cockpit) und Betroffene

In der KI-Verordnung der EU wird unter Artikel 14 „Menschliche Aufsicht“ eben jene als Bedingung für den Betrieb von Hochrisiko-Systemen festgelegt. Diese Anforderung zielt darauf ab, Risiken für Gesundheit, Sicherheit und Grundrechte der Betroffenen zu minimieren. Hieraus leiten sich neue Anforderungen an KI-Systeme ab, die in der EU zum Einsatz kommen sollen. KI-Systeme müssen so gestaltet werden, dass sie während ihrer Nutzung durch geeignete Mensch-Maschine-Schnittstellen von Menschen beaufsichtigt werden können (KI -Verordnung, Art. 14 [1]).

Das Forschungsprojekt KI-Cockpit reagiert auf diese Anforderung einer menschlichen Aufsicht und Letztentscheidung mit der Entwicklung der KI-Cockpit-Software.[6] Dieses KI-Cockpit soll das produktive Miteinander von Mensch und Maschine optimieren, indem es den technisch versierten Nutzer:innen (sogenannten KIC-Operator:innen, s. u.) erlaubt, die Entscheidungen des Systems zu verstehen, zu kontrollieren und zu beeinflussen, wie es etwa beim Autopiloten in der Luftfahrt längst etabliert ist.

Softwarehersteller haben nach der KI-Verordnung der EU die Aufgabe, entsprechende Mensch-Maschine-Schnittstellen zu implementieren. Anschließend müssten die Betreibenden laut KI-Verordnung eine bzw. mehrere Personen auswählen, die die menschliche Aufsicht operativ durchführen (Art. 26 [2]). Das KI-Cockpit wird dabei von einem/einer „*KIC-Operator:in*“ bedient, die dafür entsprechend geschult bzw. qualifiziert ist. Der Expert:innenstatus leitet sich aus den in Artikel 14(4) der KI-Verordnung formulierten Anforderungen an die menschliche Aufsicht ab. Zu den genannten Forderungen zählen neben Aspekten des technischen Fachwissens[7] auch Aspekte der kontextgebundenen Entscheidungsfindung.[8] Das heißt, die Fachperson muss die zugrunde liegenden Prozesse sowie die Möglichkeiten der Kontrolle und des Eingriffs verstehen und kennen. Entsprechend ist das KI-Cockpit als Expert:innen-Software ausgelegt.

Im Projekt wird zwischen folgenden **drei Rollen** unterschieden, die bei der Gestaltung menschlicher Aufsicht mitzudenken sind. Sie werden alle in unterschiedlicher Weise durch die entwickelten Produkte befähigt und sind in unterschiedlichem Maße von potenziellen KI-Entscheidungen betroffen (Abbildung 1):

[6] Das KI-Cockpit Projekt orientiert sich hierbei am gesetzten Rechtsrahmen, kann aber nicht rechtsverbindlich sicherstellen, dass der Einsatz eines KI-Cockpits auch zu 100 % den Anforderungen der KI-Verordnung entspricht. Dies geht auf den Umstand zurück, dass sich erst in der Rechtsanwendung zeigen wird, was als adäquater Umgang mit den KI-Risiken gilt.

[7] Dazu zählen z. B. „Fähigkeiten und Grenzen des Hochrisiko-KI-Systems angemessen verstehen und […] überwachen“ (Art. 14 (4a) und „Ausgabe des Hochrisiko-KI-Systems richtig zu interpretieren“ (Art. 14 (4c)).

[8] Dazu zählen z. B. Automatisierungs-Bias reflektieren (Art. 14 (4b)); nicht Verwendung des Hochrisiko-KI-Systems (Art. 14 (4d)); Eingriff ins System und Stopp-Taste (Art. 14 (4e)).

1) KIC-Operator:in

Darunter sind Expert:innen, z. B. in der Rolle einer Fachaufsicht, zu verstehen, die das KI-Cockpit bedienen. Diese Rolle kann beispielsweise analog zu eine:r Datenschutzbeauftragte:n zugeschnitten sein.

2) Endnutzer:innen

Diese sind Anwender:innen der KI-basierten Software, die keinen Zugriff auf das KI-Cockpit haben. Ihnen steht nach Artikel 50(1) und Artikel 26(7) der KI-Verordnung unter Umständen ein Recht zu, über den Einsatz von KI informiert zu werden. Für sie wird *menschlichen Aufsicht* zum einen durch die/den *KIC-Operator:in* gewährleistet, zum anderen werden sie, über die Vorgaben der KI-Verordnung hinausgehend, durch das Transparenz-Interface angesprochen und zu informierten Entscheidungen über die Nutzung befähigt.

3) Betroffene

Betroffene Personen interagieren nicht direkt mit dem KI-System, sind aber von den KI-Entscheidungen tangiert. Das können beispielsweise Verkehrsteilnehmende sein, die basierend auf KI-gestützten Entscheidungen am Straßenverkehr teilnehmen. Sie haben keine eigenen Kontrollmöglichkeiten und werden durch die menschliche Aufsicht der/des *KIC-Operator:in* geschützt.

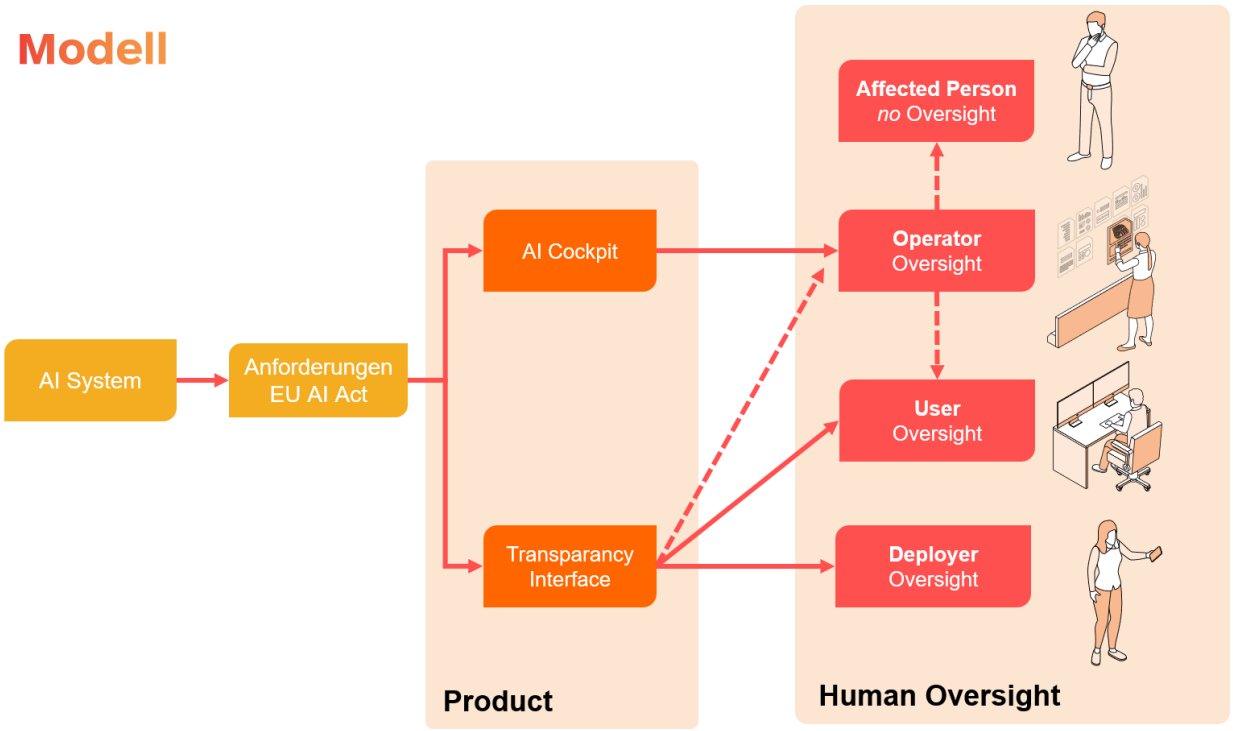

Abbildung 1: Unterscheidung der vier Nutzer:innengruppen im KI-Cockpit

2.2 Human in Command: Der Mensch im Mittelpunkt

Kurzgefasst

- Das „Human in Command"-Konzept stellt den Menschen als zentrale Instanz in der Steuerung und Beaufsichtigung von KI-Systemen in den Mittelpunkt des soziotechnischen Systems.
- Das Projekt setzt das abstrakte „Human in Command"-Konzept praktisch um, indem es den Menschen aktiv in die Steuerung und Gestaltung von KI-Systemen einbindet.
- Partizipative Technikentwicklung („Design4Command") integriert kontinuierliches Nutzer:innenfeedback aus Praxistests, um das Design an sozialen und psychologischen Bedürfnissen der Nutzer:innen auszurichten.
- Durch interdisziplinäre Ansätze aus Sozialforschung, Psychologie, Design-Engineering und Neurowissenschaften wird das KI-Cockpit entwickelt, das die Prinzipien des „Human in Command" in der Praxis realisiert.

Die Entwicklung des KI-Cockpit folgt dem „*Human in Command*"-Konzept.[9] In diesem Konzept drückt sich zentral die normative Perspektive aus, dass der Mensch im Zentrum des soziotechnischen Systems steht. In der KI-Verordnung findet sich diese normative Forderung in sprachlich abgemilderter Form als „menschliche Aufsicht" wieder. Darunter werden gemäß KI-Verordnung neben der wirksamen Beaufsichtigung durch eine natürliche Person auch aktive Eingriffe verstanden, z. B. durch das Drücken einer Stopptaste. Das Konzept „Human in Command" geht jedoch hierüber hinaus, indem es dem Menschen nicht nur bei der Vermeidung von Risiken und Schäden eine aktive Rolle und Verantwortung zuspricht, sondern auch die Potenziale der menschlichen Steuerung betont: einen verbesserten Nutzen und erhöhte Akzeptanz.

Anders als im etablierten Konzept „*Human in the Loop*" setzt „*Human in Command*" bereits in der Designphase des jeweiligen Systems an und bezieht menschliche Übersicht und Aufsicht hier bereits durch einen partizipativen Entwicklungsansatz ein. Auf diese Weise werden im KI-Cockpit-Projekt Mensch und Technik nicht getrennt voneinander, sondern als sogenanntes soziotechnisches System betrachtet.[10]
Das System wird mithilfe von Methoden der partizipativen Technikentwicklung gebildet, empirisch erprobt und wissenschaftlich begleitet. Die Ergebnisse der wissenschaftlich begleiteten Praxistests, etwa zur Akzeptanz, dem Kontrollempfinden und der Stressbelastung in der Nutzung, fließen über mehrere Feedbackschleifen zurück ins Design. Dieses iterative, partizipative und an den psychologischen und sozialen Bedürfnissen der Nutzer:innen ausgerichtete Vorgehen wird als *Design4Command* bezeichnet.

Basierend auf diesem Ansatz werden als technische Lösungen das *KI-Cockpit* und das *Transparenz-Interface* entwickelt. Das Design wird dabei aus insgesamt vier Forschungsrichtungen bereichert, die das Zusammenwirken von Mensch und Technik untersuchen sowie das notwendige Wissen und Können der Nutzer:innen evaluieren. Hierbei ergänzen sich Methoden der **qualitativen Sozialforschung** (teilnehmende Beobachtung, partizipative Designworkshops, Fokusgruppen), **Psychologie** (Studien zur Auswirkung einzelner KI-Cockpits bzw. Transparenzinterface-Merkmale auf die Akzeptanz, die Zufriedenheit und das Vertrauen in ein KI-Sys-

[9] Aschenbrenner, D., Colloseus, C. (2023). Human in Command in Manufacturing. In: Alfnes, E., Romsdal, A., Strandhagen, J.O., von Cieminski, G., Romero, D. (eds) Advances in Production Management Systems. Production Management Systems for Responsible Manufacturing, Service, and Logistics Futures. APMS 2023. IFIP Advances in Information and Communication Technology, vol 689. Springer, Cham, pp. 559-572.

[10] Sawyer, Steve & Jarrahi, Mohammad Hossein. (2015). The Sociotechnical Perspective. In: Information Systems and Information Technology, Volume 2 (Computing Handbook Set) Edition: Third Edition Publisher: Chapman and Hall/CRC, Editors: Heikki Topi and Allen Tucker

tem), **Design-Engineering** (Designmethoden, Informationsdesign, Systems Engineering) und **neurowissenschaftliche Methoden** zur Aufmerksamkeitsverteilung, um ein der Rolle der/des *KIC-Operator:in* möglichst angemessenes Design zu entwickeln und schrittweise zu verbessern.

2.3 Umsetzung der menschlichen Aufsicht im KI-Cockpit und Transparenz-Interface

2.3.1 Die Überwachungs- und Steuerungssoftware KI-Cockpit

Kurzgefasst

- Das KI-Cockpit ermöglicht die Steuerung und Überwachung von KI-Systemen durch menschliche Operator:innen, bietet Transparenz durch Performance- und Fairness-KPIs sowie eine zentrale Plattform zur Visualisierung aller relevanten Daten.
- Das KI-Cockpit bietet vier zentrale Funktionen: Performance-Monitoring, Testfälle zur Überprüfung von diskriminierungsrelevanten Merkmalen, Einzelfallanalysen und kontextspezifische Autonomiestufen, die menschliche Eingriffe ermöglichen.
- Das KI-Cockpit wird nutzerfreundlich und partizipativ gestaltet, um auch Anwender:innen ohne technisches Wissen die Kontrolle und das Verständnis von KI-Entscheidungen zu ermöglichen, und als Open-Source-Anwendung zur Verfügung gestellt.

Die beiden technischen Produkte, die basierend auf dem Ansatz *Human in Command* entwickelt werden, sind das KI-Cockpit und das Transparenz-Interface. Sie werden durch ein Vorgehensmodell ergänzt, welches ihren Einsatz anleitet, die Designentscheidungen kontextualisiert und die gesammelten Best Practices aus dem Projekt für die zukünftigen Nutzer:innen dokumentiert.

Das KI-Cockpit

Das KI-Cockpit ist ein spezifischer Ansatz zur Steuerung und Überwachung von künstlicher Intelligenz. Das Systemmodell des KI-Cockpits wird in Abbildung 2 dargestellt: Der Prozess beginnt mit dem KI-Algorithmus, der innerhalb der vorgegebenen Parameter arbeitet (0). Die generierten Daten werden dann in einem *Human-in-the-Loop*-Ansatz in das Cockpit überführt (1) und visualisiert an den Menschen weitergeleitet (2). Dieser übernimmt eine Überwachungs- und Bewertungsfunktion, indem er die Handlungen des Algorithmus kritisch einschätzt (3). Entscheidet der Mensch, dass ein Eingriff nötig ist (4), kann dieser über das Interface des Cockpits vorgenommen werden (5). Die menschlichen Eingaben werden in algorithmische Anpassungen übersetzt (6), die dann auf das Ausgangssystem zurückwirken (7). Angepasstes Verhalten kann z. B. das Überführen in eine andere Autonomiestufe sein, bei der der Mensch direkt entscheidet, oder das Retraining der KI basierend auf den getroffenen Entscheidungen in (4).

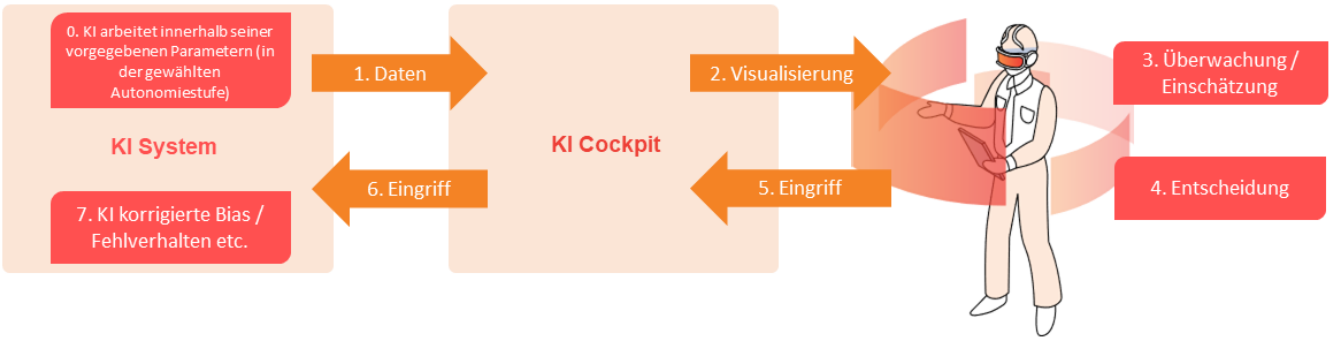

Abbildung 2: Funktionsweise des KI-Cockpits

Das entwickelte KI-Cockpit richtet sich an Hersteller und Betreibende von KI-Software. Es enthält in der aktuellen Ausführung **vier zentrale Funktionen**, mit denen das KI-Cockpit den mit der EU-KI-Verordnung konformen Betrieb der angeschlossenen KI-Systeme nach Artikel 14 ermöglichen soll. Diese Funktionen bieten umfassende Möglichkeiten zur Überwachung und Steuerung von KI-Systemen und müssen für die konkreten KI-Systeme, die durch das KI-Cockpit überwacht und gesteuert werden, konfiguriert werden.

Funktionen des KI-Cockpits

1) Performance, Fairness-KPIs (Key Performance Indicators) und Alarmierung

Hierbei handelt es sich um Indikatoren für die Performance und Fairness des Systems. Diese informieren über auditive und visuelle Hinweise und Alarme den/die *KIC-Operator:in* als Aufsichtsperson, wenn bei den festgelegten Kriterien Abweichungen bzw. Fehler erkannt werden.

2) Testfälle

Eine Schnittstelle ermöglicht das Einpflegen eigener Testfälle. Dadurch lassen sich beispielsweise einzelne diskriminierungsrelevante Merkmale ändern, um die KI-Entscheidung testweise zu durchlaufen. Dies ermöglicht wiederum die empirische Prüfung, ob derartige Merkmale einen problematischen Einfluss auf die Entscheidung ausüben.

3) Analyse von Einzelfallentscheidungen

Wo möglich, sollen Einzelfälle durch Methoden aus dem Baukasten der Erklärbaren künstlichen Intelligenz analysiert werden, die über eine Programmierschnittstelle (engl.: application programming interface, API) ins KI-Cockpit eingebunden werden und die Gründe für die Entscheidung dem/der Operator:in nachvollziehbar darstellen.

4) Autonomiestufen

Autonomiestufen ermöglichen es, steuernd in das KI-System einzugreifen. Die Autonomiestufen sind dabei kontextspezifisch festzulegen und umfassen verschiedene Ebenen: von einer Ebene, die einer Stopp-Taste gleichkommt (z. B. Parkraum-Leitung wieder manuell oder regelbasiert statt KI-gestützt; Jobmatches müssen einzeln durch *KIC-Operator:in* freigegeben werden) bis maximal zu einer vollständigen Automatisierung. Die Autonomiestufen können im KI-Cockpit festgelegt und an die oben beschriebenen Alarmierungen gekoppelt werden.

Merkmale des KI-Cockpits

In Bezug auf die effektive Einbettung der existierenden KI-Lösung stehen die **Autonomiestufen** im Mittelpunkt. Dieses in der Luftfahrt oder auch im autonomen Fahren etablierte Konzept wird im KI-Cockpit auf weitere Anwendungsfelder ausgeweitet: In unkritischen Fällen soll das System dem Menschen die Arbeit abnehmen, in kritischen Fällen allerdings nachfragen, ob es das auch machen darf. Damit bleibt die Letztentscheidung beim Menschen. Der Mensch kann auch entscheiden, dass er oder sie die volle Kontrolle ohne Assistenz durch das KI-System übernehmen möchte. Das ist meist nicht der gewünschte Zustand, da viele Fälle automatisiert bearbeitet werden sollen. Angestrebt wird daher eine klare Definition, wer was übernehmen soll. Das KI-Cockpit unterstützt diesen Vorgang. So werden beispielsweise in einer abgegrenzten Autonomie Teilbereiche festgelegt, in denen das System operieren darf und der Mensch die Lösungsvorschläge des Systems noch autorisieren muss.

Das KI-Cockpit umfasst dabei Mechanismen für eine **proaktive Überwachung und Warnung** der KIC-Operator:innen, indem es Fehler bzw. potenzielle Probleme frühzeitig erkennt und die Nutzer:innen benachrichtigt. Auf diese Weise können diese schnell Korrekturmaßnahmen ergreifen.

Das KI-Cockpit bietet zudem eine **zentrale Plattform** („Dashboard“), auf der alle relevanten Daten zur Leistung und zum Verhalten von KI-Systemen zusammengefasst und visualisiert werden. Dies ermöglicht eine ganzheitliche Sicht auf den Status und die Aktivitäten der KI. Indem die Software in den betrieblichen Ablauf integriert wird, soll sie Mitarbeitende befähigen, die Übersicht und Kontrolle über KI-Systeme zu behalten.

In diesem Zusammenhang sind Fairness- und Leistungskriterien für die Erreichung von **Transparenz und Nachvollziehbarkeit** zentral: Das KI-Cockpit-Vorgehensmodell zeigt, wie entsprechende Kriterien festgelegt und daraus Metriken entwickelt werden. Durch geeignete Visualisierungen können Nutzer:innen einschätzen, ob die Ergebnisse im Bereich des Erwartbaren und Erwünschten liegen (Situation Awareness und SIPA, Social Information Processing Awareness). So können sie nachvollziehen, wie Entscheidungen getroffen werden und warum bestimmte Ergebnisse erzielt wurden.

Zentrales Entwicklungsziel ist außerdem eine hohe **Benutzer:innenfreundlichkeit**, d. h., das Design und die Benutzer:innenoberfläche des KI-Cockpits sind auf eine einfache, intuitive Bedienbarkeit und Zugänglichkeit ausgelegt. Besonders wichtig ist dies mit Blick auf die soziotechnischen Aspekte der Technologieentwicklung. Denn die nutzer:innenzentrierte Gestaltung von KI-Anwendungen gilt als Faktor für die notwendige Entwicklung zu mehr Diversität. Hier geht es vor allem um die Frage, wie Technologie „für alle“ entwickelt werden kann. Der Fokus liegt auf einer partizipativen Technikgestaltung, bei der das Design gemeinsam mit den Nutzenden entwickelt wird. Dadurch können auch Nutzer:innen ohne tiefgehendes technisches Wissen grundlegende Aspekte der Funktionsweise und Datenverarbeitung verstehen.

Um eine niedrigschwellige Weiterentwicklung und Nachnutzung in unterschiedlichen Anwendungsfällen zu ermöglichen, wird im Rahmen des Forschungsprojekts ein allgemeingültiges interdisziplinäres Cockpit-Modell entwickelt. Es ist auf den Entwurf anderer KI-Systeme übertragbar und wird anschließend als **Open-Source-Anwendung** der Öffentlichkeit zur Verfügung gestellt.

2.3.2 Das Transparenz-Interface als Voraussetzung für selbstbestimmte Entscheidungen über das KI-System

Kurzgefasst

- Mit dem Transparenz-Interface werden kurze zentrale Informationen darüber zur Verfügung gestellt, mit welchem Ziel und Zweck und auf welche Weise das KI-System eingesetzt wird.
- Dabei stellt das Transparenz-Interface Informationen in unterschiedlichem Umfang zur Verfügung, die sich nach den individuellen Informationsbedürfnissen der Anwender:innen richten und eine selbstbestimmte Entscheidung ermöglichen.

Die in der EU-KI-Verordnung getroffene Festlegung, dass die menschliche Aufsicht fachkundigen Personen obliegt, die durch den Betreiber bestimmt werden, kann mit dem KI-Cockpit gut umgesetzt werden. Gleichzeitig bleibt jedoch ein zentraler Schritt der menschlichen Aufsicht ungeregelt: Obwohl Artikel 50 (1) und Artikel 26 (7) unter bestimmten Umständen vorsehen, dass die Betreiber die Nutzer:innen über den Einsatz von KI informieren müssen, erscheint es sinnvoll, die Transparenz für die Endnutzer:innen zu erweitern, um eine selbstbestimmte Entscheidung über den Einsatz sicherzustellen. Während der Einsatz von KI viele Potenziale bietet, gibt es ebenfalls gute Gründe, solche Systeme in spezifischen Kontexten nicht einzusetzen. Im Großen wird diese Abwägung in Artikel 13 bei den Transparenzpflichten der Herstellenden gegenüber den Betreibenden abgedeckt. Auf der Ebene der konkreten menschlichen Aufsicht sind es bei vielen KI-Werkzeugen aber doch die Endnutzer:innen, die im Einzelfall entscheiden müssen, welches der zur Verfügung stehenden Werkzeuge sie für eine Aufgabe wählen. Die sinnvolle Notwendigkeit eines *Transparenz-Interface* soll mit zwei Gründen dargelegt werden:

So mag erstens das KI-basierte Verkehrsmanagement im Alltag effizienter und genauer funktionieren; bei einer innerstädtischen Großveranstaltung (z. B. Marathon) wäre es aber gegebenenfalls besser, auf ein weniger kluges, jedoch auch weniger komplexes „klassisches“ Verkehrsmanagement zu wechseln. Ebenfalls ist es für die Pflegekraft essenziell, die Schwachstellen und Herausforderungen von KI-basierter Pflegedokumentation zu kennen, um einen Automation-Bias zu verhindern und für besonders kritische Anwendungsfälle zu sensibilisieren.

Zweitens hängt die Akzeptanz von KI-Systemen auch von ihrer Transparenz ab. Der in der Gesellschaft geführte Diskurs über KI ist auf der einen Seite von Heilsversprechen und auf der anderen Seite von Schreckensszenarien geprägt. Nicht zuletzt wurde über die breite Verfügbarkeit von Chatbots in den letzten Jahren für alle erfahrbar, was KI kann, aber auch, welche Schwächen sie hat. Es ist daher besonders wichtig, die Nutzer:innen über die Eigenschaften des Systems zu informieren. Durch Transparenz kann Vertrauen und damit Akzeptanz für die Technik geschaffen werden, welche notwendig ist, um die wirtschaftlichen und gesellschaftlichen Potenziale von KI nutzbar zu machen.

Entsprechend geht das Projekt über die Anforderungen der KI-Verordnung hinaus und erarbeitet Designvorschläge für ein *Transparenz-Interface*. Das *Transparenz-Interface* ist nicht Teil der oben beschriebenen KI-Cockpit-Software, sondern richtet sich an die Endnutzer:innen, die das KI-System einsetzen.

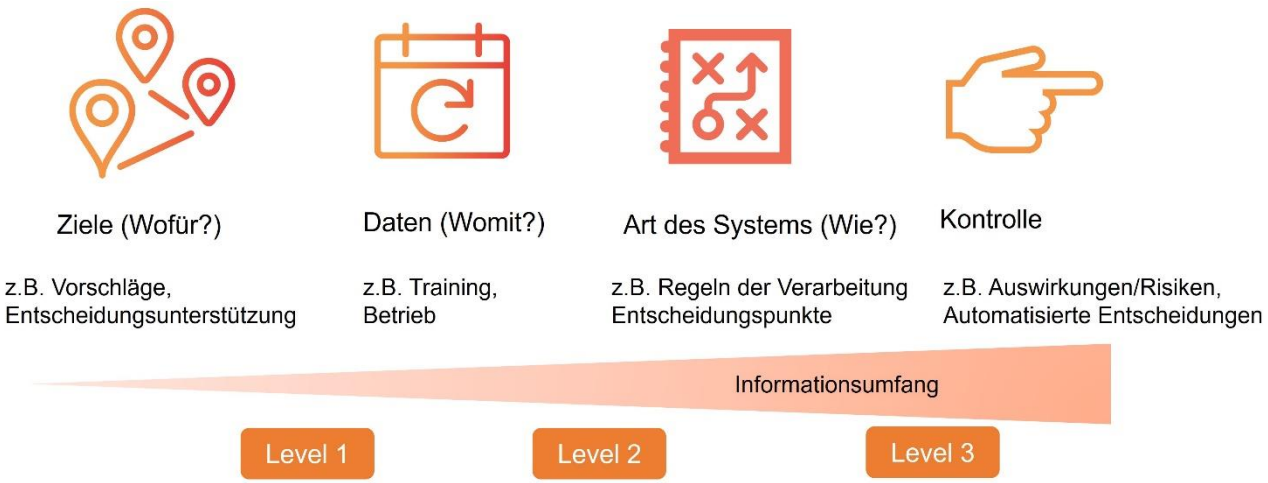

Abbildung 3: Drei Ebenen des Transparenz-Interfaces (eigene Darstellung)

Mit dem *Transparenz-Interface* werden zentrale Informationen darüber zur Verfügung gestellt, mit welchem Ziel und Zweck und auf welche Weise das KI-System eingesetzt wird. Dabei stellt das *Transparenz-Interface* Informationen in unterschiedlichem Umfang zur Verfügung. Abbildung 3 zeigt eine Übersicht über den Aufbau des *Transparenz-Interfaces*. Sie bildet die unterschiedlichen Kategorien des Informationsumfangs ab:

- Allgemeine Informationen über den **Zweck des KI-Einsatzes** (z. B. Vorauswahl von Bewerber:innen)
- Informationen über die **Art des KI-Systems** (z. B. deskriptives System, diagnostisches System, prädiktives System, präskriptives System)
- Informationen zu den **verwendeten Daten** (z. B. Informationen aus dem Lebenslauf der Bewerber:in) und
- Informationen über die **Kontrollmöglichkeiten** (z. B. Einsatz eines KI-Cockpits zur Gewährleistung menschlicher Aufsicht und Eingriffsmöglichkeit)

Die Informationen über das spezifische KI-System werden auf drei Ebenen unterschiedlichen Umfangs aufbereitet. Hierdurch soll eine zu den individuellen Informationsbedürfnissen passende Informationsplattform für die selbstbestimmte Entscheidung geschaffen werden.

2.3.3 Das Vorgehensmodell zur praktischen Anleitung und Darstellung von Best Practices

Das *KI-Cockpit* und das *Transparenz-Interface* sind die beiden Softwareprodukte, die im Projekt entwickelt werden. Diese werden flankiert durch ein *Vorgehensmodell*. Darin werden die Designentscheidungen begründet, die Implementierungen von *KI-Cockpit* und *Transparenz-Interface* angeleitet und die im Projekt gesammelten Erfahrungen in Form von Best Practices und Konfigurationsvorschlägen geteilt.

Vorgehensmodell, KI-Cockpit und Transparenz-Interface bilden zusammen die zentralen Projektergebnisse. Die Ergebnisse werden empirisch an den drei sehr unterschiedlichen Anwendungsfällen Human Resources, smarte Kommune und smarte Pflege erprobt – sie sind daher auf ein weites Anwendungsfeld übertragbar. Zudem sind sie wissenschaftlich begleitet. So fließen die Forschungsergebnisse und das fachspezifische Praxiswissen der Projektbeteiligten zusammen. Über standardisierte und nicht standardisierte Methoden erheben die forschenden Projektteilnehmenden in Zusammenarbeit mit den Fieldlabs Informationen zu unterschiedlichen Strängen, die sowohl in die Gestaltung der Technik als auch in das Vorgehensmodell einfließen.

Am Projektende werden drei praktische Implementierungen der Produkte realisiert, iterativ erprobt und verbessert vorliegen. Hierüber lassen sich konkrete Schlüsse über die Anwendungsdomänen ziehen und Transferpotenziale erschließen. Diese Prototypen und die generalisierten Varianten von KI-Cockpit und Transparenz-Interface können als Leuchttürme in die Anwendungsdomänen und darüber hinaus strahlen und Lösungen für die Herausforderungen der Durchsetzung von KI in Deutschland aufzeigen.

3. POTENZIALE FÜR ARBEITSMARKT UND SOZIALSTAAT

Kurzgefasst

- Das KI-Cockpit zeigt, wie qualitativ hochwertige, vertrauenswürdige und transparente KI-Anwendungen nach der EU-KI-Verordnung entstehen können, die Mitarbeitende entlasten und gleichzeitig die Akzeptanz für KI stärken.
- Durch den menschenzentrierten Ansatz werden KI-Potenziale genutzt, um Produktivität zu steigern und Freiräume für wertvolle Tätigkeiten wie in der Pflege oder strategische Aufgaben zu schaffen.
- Transparenz, Partizipation und Kompetenzförderung sind entscheidend für den erfolgreichen Einsatz von KI, während Risiken wie Arbeitsverdichtung durch frühzeitige Einbindung von Beschäftigten und sozialpartnerschaftlichen Akteuren minimiert werden müssen.

Das KI-Cockpit zeigt beispielhaft Wege auf, wie durch die Umsetzung der EU-KI-Verordnung qualitativ hochwertige, vertrauenswürdige und transparente KI-Anwendungen entstehen können, die Menschen im Arbeitsalltag entlasten. Mit dem „Hybrid Intelligence"-Ansatz werden die Bedürfnisse der Mitarbeitenden ins Zentrum gestellt, um das Verbesserungspotenzial von KI-Innovationen ganz auszuschöpfen. Das Projekt schafft die Infrastruktur, die nachhaltig die Akzeptanz von KI stärkt und damit ein Alleinstellungsmerkmal innerhalb des deutschen und europäischen KI-Diskurses bildet. Dabei zeichnen sich verschiedene Potenziale für Arbeitsmarkt und Sozialstaat ab.

Transformation der Arbeitswelt – aber menschenzentriert

Die verstärkte Einführung und Verbreitung von KI wird in naher und mittlerer Zukunft voraussichtlich zu einer Umstrukturierung der Beschäftigung und von Berufen führen. Durch den Einsatz dieser Technologie können noch mehr Arbeitsprozesse als bisher (teil-)automatisiert werden. KI ist in der Lage, sowohl repetitive als auch generative, ressourcenintensive Aufgaben wie Konstruktion, Programmierung und Medienerstellung teilweise zu übernehmen. Speziell regelbasierte Tätigkeiten kann KI sehr viel effizienter erledigen als der Mensch. Gleichzeitig kann die Technologie insbesondere im Hochrisikobereich (z. B. Personalwesen) nicht alleingelassen werden. Das Transparenz-Interface als vorab verständlich zur Verfügung gestellte Information und das KI-Cockpit als Schnittstelle zwischen Menschen und KI ermöglichen es, technologische Effizienzpotenziale zu heben, ohne dass Beschäftigte Kontrolle, Verantwortung und Selbstwirksamkeitsgefühl an KI-basierte Maschinen abgeben (müssen).

Wenn KI-Systeme künftig verstärkt Tätigkeiten übernehmen, die heute noch von Menschen verrichtet werden, könnte dies zu steigender Produktivität führen. Löst sich dieses Versprechen ein, entstünden Freiräume für Beschäftigte, die z. B. anstelle repetitiver Aufgaben ihre Zeit noch sinnbringender einsetzen könnten, in Tätigkeiten, die auch in Zukunft nicht von Maschinen übernommen werden können. Eine erfolgreiche Umsetzung des KI-Cockpits beispielsweise im Anwendungsfeld Pflege würde dazu führen, dass freie Kapazitäten für die besonders wertvolle Arbeit eingesetzt werden können: die persönliche Zuwendung hin zu den Patient:innen. Zur Entlastung der Verwaltung sind Automatisierungsprozesse auch im Verkehrswesen notwendig: So können mittels Datenerfassung beispielsweise in der Parkraumbewirtschaftung und Verkehrslenkung Staus und Unfälle vermieden und folglich eine effizientere Mobilität ermöglicht werden.[11] Insgesamt ergeben sich für Beschäftigte neue Freiräume für die berufliche Selbstverwirklichung und Weiterentwicklung (Lebenslanges Lernen), insbesondere durch die Integration effektiver

[11] Erste Potenziale zur Automatisierung im Verkehrswesen wurden von der Stadtverwaltung Wolfsburg im Rahmen des Forschungsprojekts positiv bewertet.

Qualifizierungsformate in den Arbeitsalltag (Lernen im Prozess). Auch können die neu geschaffenen Kapazitäten verstärkt in strategische und kreative Gestaltungsaufgaben investiert werden. Denn durch die Implementierung verschiedener KI-Technologien werden perspektivisch neue Tätigkeitsprofile mit höheren Qualifikationsanforderungen entstehen. So setzt etwa die neue Rolle des/der KIC-Operator:in technische Kompetenzen voraus, die über ein Grundverständnis von KI-Systemen hinausgehen. Ein Grundverständnis für KI ermöglicht zudem einen offenen gesellschaftlichen Diskurs zur KI-Technologie, der gleichzeitig positiven Einfluss auf die Akzeptanz und das Vertrauen der Gesellschaft in KI-Systeme hat.

Transparenz, Partizipation und Kompetenzförderung bilden Grundlage für Erfolg

Der Einsatz von KI ermöglicht es Unternehmen, ihre Produktivität deutlich zu steigern und höhere Unternehmensgewinne zu erzielen. Beispielsweise können im Bereich Personalwesen sowohl die Personalvermittlung selbst als auch ihre Kund:innen von passgenauen Matchingprozessen enorm profitieren. Erstere arbeiten deutlich effizienter, während letztere offene Stellen noch schneller und treffsicherer besetzen können. Solche Entwicklungen können volkswirtschaftlich Spielräume bei der Entlohnung schaffen. Dies schließt auch Geringqualifizierte ein, die durch eine entsprechende Erhöhung des Lohnniveaus ebenfalls von einem gesicherten Wachstum profitieren. Darüber hinaus haben Beschäftigte die Chance, sich beruflich weiterzuentwickeln, wenn sie lernen, KI-Anwendungen sicher und effektiv zu nutzen.[12] Um die Interaktion mit KI-Systemen kontrolliert gestalten zu können, sind arbeitsintegrierte Lernprozesse zentral. Die frühzeitige Kompetenzentwicklung der Mitarbeitenden durch die Unternehmen ist daher eine unverzichtbare strategische Investition, die angesichts des Fachkräftemangels insbesondere im IKT-Bereich noch stärker politisch flankiert werden muss.

Damit Beschäftigte KI-Systeme akzeptieren und deren Umsetzung proaktiv mitgestalten, bedarf es zudem größtmöglicher Transparenz sowie der Einbindung sozialpartnerschaftlicher Akteure, etwa im Rahmen betrieblicher Beteiligungsprozesse. Hilfreich ist es dabei, Beschäftigte frühzeitig an die neuen Technologien heranzuführen und sie mit ihrer Fachexpertise systematisch in die Entwicklung von KI sowie in die betriebliche Umsetzung einzubeziehen.[13] Im Rahmen des verfolgten „Human in Command“-Ansatzes schafft das Projekt KI-Cockpit hierfür sehr gute Voraussetzungen − insbesondere durch das *Transparenz-Interface*. Damit können sich Endnutzer:innen eigenständig über die KI-gestützte Verarbeitung ihrer Daten informieren. In diesem Sinne geht das *Transparenz-Interface* über die in der KI-Verordnung festgelegten Transparenzstandards hinaus. Durch diese gezielte Partizipation der Beschäftigten gelingt der erfolgreiche Einsatz praxisnaher und nutzbringender Anwendungen, die Belegschaft wird durch mehr „gute Arbeit“[14] gestärkt und das Betriebsklima insgesamt verbessert: Die Beschäftigten fühlen sich als zentrale Akteure der Organisationsentwicklung anerkannt und wertgeschätzt − und nicht durch KI-Systeme bedroht.

Bleibende Risiken der Digitalisierung und Automatisierung

Bei allem Grund zur Zuversicht hinsichtlich der vielfältigen positiven Auswirkungen von menschenzentrierter KI auf die Arbeitswelt darf nicht übersehen werden, dass die Einführung von KI auch Risiken mit sich bringen kann. So zeigt der Transformationsprozess der Digitalisierung, dass die zunehmende Implementierung digitaler Technologien und Automatisierung sowohl zu Effizienzsteigerungen als auch zu einer Verdichtung

[12] Vgl. Chen, N., Li, Z., & Tang, B. (2022). Can digital skill protect against job dis-placement risk caused by artificial intelligence? Empirical evidence from 701 detailed occupations. PLoS One, 17(11), e0277280.

[13] Vgl. Burchardt, A., Aschenbrenner, D. (2024). Praxisleitfaden KI = Kollaborativ und Interdisziplinär. In I. Knappertsbusch (Hrsg.), K. Gondlach (Hrsg.) Arbeitswelt und KI 2030 - Herausforderungen und Strategien für die Arbeit von morgen (2. Aufl.). Springer-Verlag GmbH. (aktualisierte Fassung); Aschenbrenner, D. (2022). Die Bedeutung von Participatory Design für Anwendungen mit künstlicher Intelligenz. In Greiner, R., Berger, D., & Böck, M. (Hrsg.) Analytics und Artificial Intelligence (pp. 252-259). Springer Gabler, Wiesbaden.

[14] Schröder, Lothar; Urban, Hans-Jürgen (2018): Gute Arbeit. Ökologie der Arbeit – Impulse für einen nachhaltigen Umbau, Bund-Verlag, Frankfurt.

der bisherigen Arbeitszeit führen kann.[15] Damit Mitarbeitende nicht nur immer mehr und komplexere Aufgaben in kürzerer Zeit erledigen müssen, was zu höherem Stress und psychischen Belastungen führt, muss die Beteiligung der Beschäftigten und ihrer Interessenvertretungen bei der Gestaltung und Einführung von KI in Unternehmen sichergestellt werden.

[15] Vgl. https://index-gute-arbeit.dgb.de/++co++e9c777a4-507f-11ed-9da8-001a4a160123..

4. GESTALTUNGSSPIELRÄUME FÜR POLITIK

Kurzgefasst

- Die nationale Politik hat durch die EU-KI-Verordnung die Möglichkeit, über die Regulierung hinaus zusätzliche Schutzmechanismen für Beschäftigte einzuführen, um Arbeitsverdichtung und Automatisierungsdruck zu minimieren.
- Förderprogramme und steuerliche Anreize können genutzt werden, um die Entwicklung und Einführung von KI-Systemen zu unterstützen, die Transparenz, Fairness und Kontrolle durch Beschäftigte gewährleisten.
- Ein politischer Fokus auf Hybrid Intelligence sowie auf AI Literacy ist notwendig, um die Kompetenzen der Beschäftigten im Umgang mit KI zu stärken und eine breite Nutzung der Technologien zu ermöglichen.
- Standardisierungsprozesse und öffentliche Beschaffung bieten der Politik Gestaltungsspielräume, um Transparenz und Kontrollmöglichkeiten in KI-Systemen zu fördern und menschenzentrierte Technologien voranzutreiben.
- Durch die Einbindung sozialpartnerschaftlicher Akteure und die Förderung von Responsible Research and Innovation (RRI) kann die Politik sicherstellen, dass KI ethisch und gesellschaftlich verantwortungsbewusst entwickelt und genutzt wird.

Mit der KI-Verordnung der EU ergeben sich bereits mit deren Inkrafttreten für die Arbeitswelt konkrete Anforderungen. Hinzu kommt, dass der nationale Gesetzgeber in Art. 2 Abs. 11 KI-VO mit der Öffnungsklausel umfassende Gestaltungsspielräume an die Hand bekommt, weitergehende Regelungen zu treffen, die Beschäftigte wirksam vor Risiken im Zusammenhang mit dem Einsatz von KI schützen.[16] Für Hochrisiko-Anwendungen wie Human Resources gilt dabei bereits jetzt: Der Mensch muss in der Lage sein, das technische System zu überwachen und steuernd einzugreifen. Obwohl die KI-Verordnung einen regulatorischen Rahmen insbesondere für das Inverkehrbringen von KI schafft, steht die Umsetzung dieser Vorgaben in konkrete technische Spezifikationen und praktikable Anwendungsmodelle noch am Anfang.

Technische Antworten auf Anforderungen der KI-Verordnung geben

Derzeit fehlt es an technischen Lösungen, die eine zuverlässige menschliche Letztentscheidung in der Interaktion mit KI-Systemen ermöglichen. Vor allem mangelt es an branchenübergreifenden Ansätzen, die für verschiedene Anwendungsbereiche adaptierbar sind. Das ist eine erhebliche Herausforderung, weil der Erfolg der Regulierung stark von der praktischen Umsetzbarkeit der geforderten menschlichen Aufsicht abhängt. Deshalb leistet das Bundesministerium für Arbeit und Soziales mit der Förderung des Projekts KI-Cockpit einen wichtigen Beitrag, damit die technischen Voraussetzungen für eine geeignete und rechtskonforme Mensch-Maschine-Schnittstelle Realität werden. Das Projekt arbeitet daran, die praktische Umsetzung von *Human Oversight* gemäß der europäischen KI-Verordnung zu untersuchen und zu demonstrieren. Im Ergebnis soll eine Bedienoberfläche geschaffen werden, die Systemabläufe transparent und kontrollierbar macht, und somit die Basis schafft, die Technologie verantwortungsbewusst und effektiv zu nutzen. Diese Ergebnisse werden gleichzeitig in die DIN-Normung (DIN/DKE NA 043-01-A2 GA „Künstliche Intelligenz“) sowie über die EU-Initiative AI Pact in die aktuelle Diskussion eingebracht.

Durch die konsequente Ausrichtung auf die menschliche Interaktion mit KI-Systemen in der Entwicklung des KI-Cockpits und von Beginn ihrer Einführung in Organisationen können Risiken minimiert und die Vertrauenswürdigkeit sowie die Passfähigkeit und Akzeptanz von KI-Technologien gestärkt werden. Die Entwicklung und Einführung des KI-Cockpits könnte somit einen wesentlichen Beitrag zur sicheren und ethisch

[16] https://www.bundestag.de/resource/blob/1002482/d6e9283aa27bf415bcf0b0f0dcfce228/Suchy.pdf

verantwortungsvollen Nutzung von KI in Europa leisten. Ziel muss es sein, dass der Mensch bestimmt, welche Aufgaben und Entscheidungen er bzw. sie übernimmt und welche das technische System erledigt. Die technologische Umsetzung als Open-Source-Lösung ermöglicht die Ausweitung des Grundgedankens auf weitere konkrete Anwendungen über die im Projekt konkret betrachteten hinaus. Als Ergänzung zu KI-Algorithmen, die ihre Entscheidungen erklären können (*Explainable AI*), kann der hier angestrebte „*Human in Command*“ -Ansatz nahezu für beliebige Anwendungen verwendet.

Menschenzentrierte Innovationen einer „Hybriden Intelligenz“ stärken

Die EU-KI-Verordnung definiert grundlegende Anforderungen, die den Menschen in der Interaktion mit KI-Technologien stärken. Um das volle ökonomische und soziale Potenzial von KI-Technologien zu nutzen, muss es darum gehen, menschenzentrierte Technologiegestaltung weiter zu denken. Es gilt nicht nur, Menschen vor den Risiken von KI-Systemen zu schützen, sondern darum, ihre Potenziale zu maximieren. Aktuell nimmt die Mehrheit der Beschäftigten die Interaktion mit KI-Systemen noch nicht als eine Neugestaltung ihrer Handlungsspielräume wahr.[17] Damit die Einführung von KI Handlungsspielräume für Beschäftigte erweitert und nicht beschränkt, ist die Beteiligung der Mitarbeitenden bei der Evaluierung und Einführung solcher Systeme essenziell. Die Fähigkeit von KI-Systemen, natürliche Sprache zu verarbeiten und dadurch große Datenmengen und Rechenoperationen niedrigschwellig zugänglich zu machen, hebt die Mensch-Technik-Interaktion auf eine neue Ebene. Um effiziente Anwendungskontexte zu finden und die Ergebnisse interaktiver Systeme richtig einzuschätzen, sind analog zu einer *Digital Literacy* oder *Sustainable Literacy* auch eine *AI Literacy* sowie die Einführung und Entwicklung mit dem Menschen im Mittelpunkt der Interaktionsgestaltung entscheidende Voraussetzungen.[18] Dabei geht es neben dem risikominimierenden Ansatz auch um einen chancenmaximierenden Ansatz: Hybride Intelligenz als gelungene Kombination einer Mensch-KI-Interaktion kann (und sollte) mehr als nur die Summe von Arbeitsleistung sein.

Standardisierung als Schlüssel zur Skalierung menschenzentrierter Designprinzipien

Die Best Practices für das Design von KI-Anwendungen kommen aktuell vor allem von den amerikanischen Hyperscalern. Durch die EU-KI-Verordnung machen sich viele verschiedene Unternehmen gleichzeitig mit Forschungs- und Entwicklungspartnern auf die Suche nach praktikablen Designmethoden für ihre Bedienoberflächen. Neben der Stärkung der menschenzentrierten Designs an sich bringt dies auch die Chance mit sich, die aus der Forschung bekannten Prinzipien und Methoden trotz damit verbundenen Aufwands und Kosten in einer schnelllebigen Industrie zu verankern. Damit Bedienoberflächen, die eine derartige Transparenz und Kontrolle ermöglichen, branchenübergreifend adaptiert werden können, bedarf es im Rahmen der Standardisierung auf Europaebene auch klarer Vorgaben − und zwar im Schulterschluss mit einflussreichen Verbänden und Interessensvertretungen: Zum einen zur Implementierung von Transparenzelementen, die eine Informationsbasis für das Eingreifen in die Systemabläufe bieten, und zum anderen von Bedienelementen, die es erlauben, kritische Abläufe des Systems stufenweise stärker manuell zu kontrollieren (Autonomiestufen).

Soziotechnische Kompetenzprofile entwickeln

Wenn Prototypen wie das KI-Cockpit und das Transparenz-Interface vorliegen und einheitliche Standards geschaffen sind, führt dies noch nicht zwangsläufig zu einer reibungslosen Adaption entsprechender

[17] Peters et. al. (2023): Arbeiten mit Künstlicher Intelligenz – fünf Kurzszenarien zur „Mensch-Technik-Interaktion 2030“, Denkfabrik Digitale Arbeitsgesellschaft, Bundesministerium für Arbeit und Soziales.

[18] Vgl. Wienrich et. al (2022): AI Literacy: Kompetenzdimensionen und Einflussfaktoren im Kontext von Arbeit. Working Paper im Rahmen des „Mensch-Technik-Interaktion“-Fachdialogs des KI-Observatoriums des BMAS, online abrufbar unter: https://www.denkfabrik-bmas.de/fileadmin/Downloads/Publikationen/AI_Literacy_Kompetenzdimensionen_und_Einflussfaktoren_im_Kontext_von_Arbeit.pdf

Lösungen. Hierzu müssen KI-anwendende Unternehmen in ihren Kompetenzen und finanziellen Ressourcen gestärkt werden. Für einen erfolgreichen Transfer menschenzentrierter Technologien benötigen Unternehmen mehr als technische Kompetenzen, z. B. Programmierkenntnisse, und domänenspezifische Kompetenzen wie den Umgang mit branchenspezifischen Softwareapplikationen und branchenspezifisches Kontextwissen. Vielmehr sind in der Unternehmenspraxis verstärkt dezidiert soziotechnische Kompetenzen vonnöten. Die Herausforderung liegt hierin, dass diese Kompetenzen bislang in technischen Studiengängen, in gewerblich-technischen Ausbildungsberufen und in einschlägigen Stellenprofilen von KI-anwendenden und KI-entwickelnden Unternehmen kaum eine Rolle spielen. Im KI-Hype darf es nicht noch einmal an der menschenzentrierten Perspektive mangeln, damit die Begeisterung über KI nicht wieder in sich zusammenfällt. Daher bedarf es einer umfassenden Stärkung von Einordnungs- und Beurteilungskompetenz bezogen auf das Zusammenspiel von Menschen, Technologie und Organisation.[19]

Förderprogramme und Abschreibungsmöglichkeiten gezielt nutzen

Damit menschenzentrierte Innovationen im KI-Bereich wie das KI-Cockpit und das Transparenz-Interface verstärkt entwickelt und adaptiert werden, erscheint es sinnvoll, bestehende Instrumente der Innovationspolitik weiterzuentwickeln. Es muss darum gehen, dass entsprechende Lösungen nicht nur in Leuchtturmprojekten in Einzelförderung entstehen. So könnte die steuerliche Forschungsförderung noch gezielter genutzt werden, um bei dezidiert soziotechnisch ausgelegten Innovationsvorhaben attraktivere Abschreibungsmöglichkeiten zu schaffen. Ein weiterer Beitrag könnte vor allem für kleine und mittlere Unternehmen die Stärkung soziotechnischer Transferberatung sein. Denkbar wäre etwa die Schaffung eigenständiger Beratungsangebote, z. B. nach dem Vorbild der Mittelstand-4.0-Kompetenzzentren, die gezielt soziotechnische Kompetenzen vermitteln und bereitstellen, um Unternehmen bei der Entwicklung menschenzentrierter Technologien zu stärken. Auch die Darstellung eines ökonomischen Mehrwerts soziotechnischer Ansätze (größere Nutzer:innenakzeptanz, langlebiger, Hype-resilientere Geschäftsmodelle etc.) müssen betrachtet und zusammen mit Unternehmen immer wieder bestätigt werden. Entgegen der technologiezentrischen Innovationsförderung muss ein gesamtwirksamer Ansatz (*Responsible Research and Innovation*) gewählt werden, um eindimensionalen Technologieszenarien entgegenzutreten (vgl. *Civic Coding).*[20]

Innovative Anbieter durch öffentliche Beschaffung stärken

Der Staat kann in diesem Zusammenhang sowohl über die Schaffung geeigneter Rahmenbedingungen als auch in einer aktiven Rolle als Nachfragender digitaler Technologien gezielt menschenzentrierte Ansätze stärken, wie sie das KI-Cockpit und das Transparenz-Interface repräsentieren. Durch das Mittel der öffentlichen Beschaffung kann der öffentliche Sektor gezielt für solche Technologien Nachfrage kreieren und den Markthochlauf unterstützen und somit die Menschen in den Mittelpunkt stellen. Ein solches Vorgehen kann dabei einen mehrfachen Vorteil schaffen: Erstens setzt der Staat so marktwirtschaftliche Anreize für die Stärkung von Angeboten menschenzentrierter Innovationen. Zweitens profitieren Beschäftigte des öffentlichen Sektors und betroffene Bürger:innen so von den Vorteilen technischer Lösungen, die ihre Informations- und Eingriffsmöglichkeiten stärken.

[19] Vgl. Hirsch-Kreinsen, H. (2018). Einleitung: Digitalisierung industrieller Arbeit. In H. Hirsch-Kreinsen, P. Ittermann & J. Niehaus (Hrsg.). Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen.

[20] „*Civic Coding* – Innovationsnetz KI für das Gemeinwohl“ ist eine gemeinsame Initiative des Bundesministeriums für Arbeit und Soziales (BMAS), des Bundesministeriums für Familie, Senioren, Frauen und Jugend (BMFSFJ) und des Bundesministeriums für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz (BMUV), mit der ressortübergreifend ein Netzwerk für gemeinwohlorientierten KI geschaffen wird.

5. LITERATURVERZEICHNIS

Akata, Zeynep et al. (2020): A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. In: Computer, vol. 53, no. 8, doi: 10.1109/MC.2020.2996587. Pp. 18 – 28.

Aschenbrenner, Doris (2022): Die Bedeutung von Participatory Design für Anwendungen mit künstlicher Intelligenz. In: Greiner, Ramona; Berger, David; Böck, Matthias: Analytics und Artificial Intelligence. Springer Gabler, Wiesbaden. Pp. 252 – 259.

Aschenbrenner, Doris; Colloseus, Cecilia (2023): Human in Command in Manufacturing. In: Alfnes, Erlend et al.: Advances in Production Management Systems. Production Management Systems for Responsible Manufacturing, Service, and Logistics Futures. Cham: Springer, pp. 559 – 572.

Bundesministerium für Arbeit und Soziales (2022): Arbeiten mit Künstlicher Intelligenz. Perspektiven für eine menschenzentrierte Gestaltung von KI. Online verfügbar unter: https://www.denkfabrik-bmas.de/fileadmin/Downloads/Publikationen/Arbeiten_mit_Kuenstlicher_Intelligenz_bf.pdf, zuletzt geprüft am 30.10.2024.

Burchardt, Aljoscha; Aschenbrenner, Doris (2024). Praxisleitfaden KI = Kollaborativ und Interdisziplinär. Verantwortungsvolle Innovation für die Integration von Anwendungen der künstlichen Intelligenz in die Arbeitswelt. In: Knappertsbusch, Inka; Gondlach, Kai: Arbeitswelt und KI 2030. Herausforderungen und Strategien für die Arbeit von morgen (2. Aufl.). Springer-Verlag GmbH (aktualisierte Fassung).

Chen, Ni; Li, Zhi; Tang, Bo (2022): Can digital skill protect against job dis-placement risk caused by artificial intelligence? Empirical evidence from 701 detailed occupations. In: PLoS One, 17(11), e0277280.

Civic Coding (2024): Forschungsbericht. Online verfügbar unter: https://www.civic-coding.de/angebote/publikationen/forschungsbericht, zuletzt geprüft am 30.10.2024.

Deutscher Bundestag (2024): Stellungnahme – Öffentliche Anhörung „Nationale Spielräume bei der Umsetzung des europäischen Gesetzes über Künstliche Intelligenz“. Ausschuss für Digitales. Online verfügbar unter: https://www.bundestag.de/resource/blob/1002482/d6e9283aa27bf415bcf0b0f0dcfce228/Suchy.pdf, zuletzt geprüft am 30.10.2024.

Deutscher Gewerkschaftsbund (2022): Report 2022: Digitale Transformation der Arbeitswelt – Veränderungen der Arbeit aus Sicht der Beschäftigten. Ergebnisse des DGB-Index Gute Arbeit 2022. Online verfügbar unter: https://index-gute-arbeit.dgb.de/++co++e9c777a4-507f-11ed-9da8-001a4a160123, zuletzt geprüft am 30.10.2024.

European Commission (2024): AI Pact. Online verfügbar unter: https://digital-strategy.ec.europa.eu/en/policies/ai-pact, zuletzt geprüft am 30.10.2024.

Hirsch-Kreinsen, Hartmut (2018). Einleitung: Digitalisierung industrieller Arbeit. In: Hirsch-Kreinsen, Hartmut; Ittermann, Peter; Niehaus, Jonathan: Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen. Nomos, 2. Auflage.

Peters, Robert; Burmeister, Klaus; Apt, Wenke (2023): Arbeiten mit Künstlicher Intelligenz – fünf Kurzszenarien zur „Mensch-Technik-Interaktion 2030“. Denkfabrik Digitale Arbeitsgesellschaft, Bundesministerium für Arbeit und Soziales. Online verfügbar unter: https://www.denkfabrik-bmas.de/fileadmin/Downloads/Publikationen/Arbeiten_mit_KI_fuenf_Szenarien_2030_bf.pdf, zuletzt geprüft am 30.10.2024.

Sawyer, Steve; Jarrahi, Mohammad Hossein (2015): The Sociotechnical Perspective. In: Topi, Heikki; Tucker, Allen: Information Systems and Information Technology, Volume 2 (Computing Handbook Set), Third Edition, Publisher: Chapman and Hall/CRC.

Schröder, Lothar; Urban, Hans-Jürgen (2018): Gute Arbeit. Ökologie der Arbeit – Impulse für einen nachhaltigen Umbau, Bund-Verlag, Frankfurt.

Wienrich, Carolin et. al (2022): AI Literacy: Kompetenzdimensionen und Einflussfaktoren im Kontext von Arbeit. Working Paper im Rahmen des „Mensch-Technik-Interaktion“-Fachdialogs des KI-Observatoriums des Bundesministeriums für Arbeit und Soziales. Online verfügbar unter: https://www.denkfabrik-bmas.de/fileadmin/Downloads/Publikationen/AI_Literacy_Kompetenzdimensionen_und_Einflussfaktoren_im_Kontext_von_Arbeit.pdf, zuletzt geprüft am 30.10.2024.

6. ABBILDUNGSVERZEICHNIS


cockpit