

Prospektive menschenzentrierte Bewertung des KI-Einsatzes als Teil einer erfolgreichen Implementationsstrategie

Martin SCHMAUDER¹, Gritt OTT², Daniel GRÖLLICH¹, Edgar SCHERSTJANOI¹,
Solveig HAUSMANN², Sylvia FRANKE-JORDAN², Sophie BERGER²

¹ Professur für Arbeitswissenschaft, TU Dresden,
Marschnerstr. 39, D-01307 Dresden

² CIMTT Zentrum für Produktionstechnik und Organisation, TU Dresden,
Helmholtzstraße 10, D-01062 Dresden

Kurzfassung: Eine erfolgreiche Implementation von KI-Lösungen erfordert ein systematisches Vorgehen, um alle relevanten Gestaltungsaspekte und deren Abhängigkeiten voneinander angemessen zu berücksichtigen. Hierzu werden Methoden aus dem Feld der Technikfolgenabschätzung aufgegriffen. Im Fokus stehen die prospektive Identifikation und Beurteilung von ethisch bzw. arbeitswissenschaftlich problematischen Auswirkungen und die Ableitung von Handlungsbedarf. Im Beitrag wird die entwickelte Vorgehensweise erläutert.

Schlüsselwörter: menschenzentrierter Bewertungsprozess, KI-Anwendungen, ethische Werte, Technikfolgenabschätzung, partizipative Reflexion, Use Cases, Personas

1. Motivation

KI-Einsatz in Wertschöpfungsprozessen ist mit Chancen und Risiken für diese Prozesse bzw. für die Arbeitskräfte verbunden, wie Digitalisierung insgesamt (Grunwald 2019). Studien zeigen Defizite im Entwicklungs-/Implementationsprozess der KI-Anwendungen, z. B. bei Auswahl der Datengrundlage, Modellerstellung und Trainieren von Modellen, woraus Risiken im Einsatz erwachsen können (Wittenbrink et al. 2023). Im Folgenden werden nicht die Risiken der Entscheidungen einer KI-Anwendung, sondern die Risiken im Umgang mit der KI-Anwendung fokussiert.

Eine erfolgreiche Implementation von KI-Lösungen erfordert ein systematisches Vorgehen, um alle relevanten Gestaltungsaspekte (Mensch – Technik – Organisation) und ihre Abhängigkeiten voneinander zu berücksichtigen. Dafür ist eine Vorgehensweise entwickelt und erprobt worden, die folgenden Anforderungen genügen soll:

- *möglichst frühzeitiger Bewertungsstart trotz Collingridge-Dilemma*

Einerseits sind Wirkungen einer Lösung nicht genau vorhersagbar, solange sich die Lösung noch in der Entwicklung befindet, andererseits sind Änderungen schwierig, je weiter fortgeschritten der Entwicklungsprozess ist (Collingridge 1980). Es bedarf also eines bewussten Umgangs mit unvollständigen bzw. ungenauen Informationen.

- *Einbindung möglichst aller relevanten Akteure und ihre Aktivierung*

Mit der Vorgehensweise sollen sowohl Unternehmensvertreter als auch Begleitforscher aus verschiedenen Arbeitsbereichen und Disziplinen einbezogen werden, um der Komplexität der Problematik zu genügen.

- **Erweiterung der arbeitswissenschaftlichen um die ethische Perspektive**

Aus arbeitswissenschaftlicher Sicht stehen die Auswirkungen auf die Arbeitstätigkeit, resultierende Veränderungen von Arbeitsbedingungen, organisatorische Abläufe und Kompetenzanforderungen im Fokus. Darüber hinaus gewinnt die Beurteilung der Eigenschaften von KI-Lösungen aus ethischer Sicht an Bedeutung. Das setzt ein Normativ bzw. gesellschaftlich akzeptierten Werte für KI-Anwendungen voraus.

- **Fallspezifische Perspektive**

Eine besondere Herausforderung ist die immense Unternehmensspezifik von KI-Lösungen. Eine generalisierte Formulierung von Gestaltungsempfehlungen zur technischen Lösung, wie bisher für Arbeitsmittel üblich, ist daher nicht möglich. Stattdessen muss fallspezifisch gearbeitet werden, um handlungsleitende Empfehlungen aussprechen zu können.

2. Erläuterung der Vorgehensweise

Die vorgeschlagene Vorgehensweise dient dazu, die verantwortlichen Akteure für die Konsequenzen der KI-Einführung zu sensibilisieren und das Bewusstsein für die resultierenden Veränderungen zu schaffen. Problematische Auswirkungen sollen partizipativ und interdisziplinär identifiziert und handlungsleitende Maßnahmen zu ihrer Reduzierung initiiert werden. Abbildung 1 zeigt im Überblick die Abschnitte des Bewertungsprozesses. Die Vorgehensweise geht vom empfohlenen Vorgehen der Technikbewertung nach VDI 3780 aus und spezifiziert die einzelnen Phasen um Arbeitsschritte, die auf die prospektive Bewertung von KI-Anwendungen ausgerichtet sind. Zur Umsetzung werden neben gesicherten arbeitswissenschaftlichen Erkenntnissen zur Bewertung Methoden aus dem Inventar der Technikfolgenabschätzung (vgl. Grundwald (2002)) aufgegriffen und an die Problemstellung angepasst.

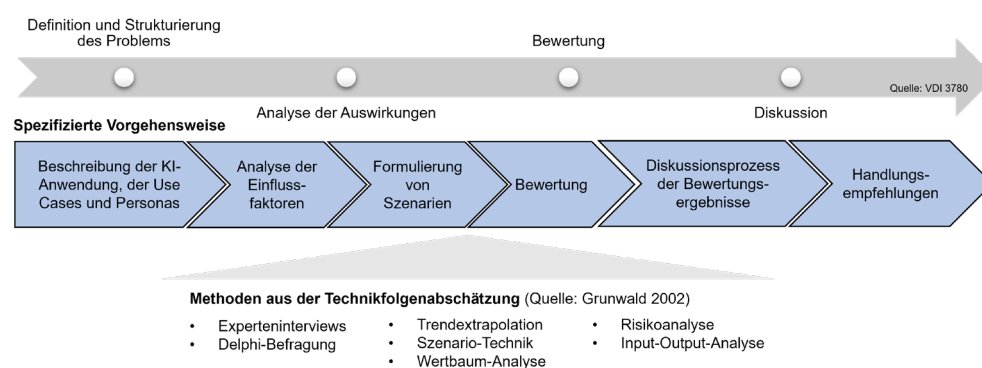


Abbildung 1: Vorgehensweise zur prospektiven menschenzentrierten Bewertung von KI-Anwendungen

Für die prospektive Bewertung wird die Methode der Szenarioanalyse (Kosow et al. 2008, Kosow et al. 2015) auf die in Kombination mit KI-Lösungen entstehenden Arbeitssysteme angewandt. Dazu wurde ein erweiterter Arbeitssystembegriff entwickelt (siehe Abbildung 2). Die Szenarioanalyse erlaubt in ihrer Kommunikationsfunktion nicht nur eine Experten- und Stakeholdereinbindung, sondern trägt zur Beherrschung unsicherer, komplexer und ambivalenter zukünftiger Entwicklungen bei (Kosow et al. 2015).

Gegenstand der Bewertung sind die resultierenden Arbeitstätigkeiten, die als Use Cases abgebildet werden. Diese werden in Beziehung zur ausführenden Person gebracht, für deren prospektive Charakterisierung das Konzept der Personas adaptiert wird.

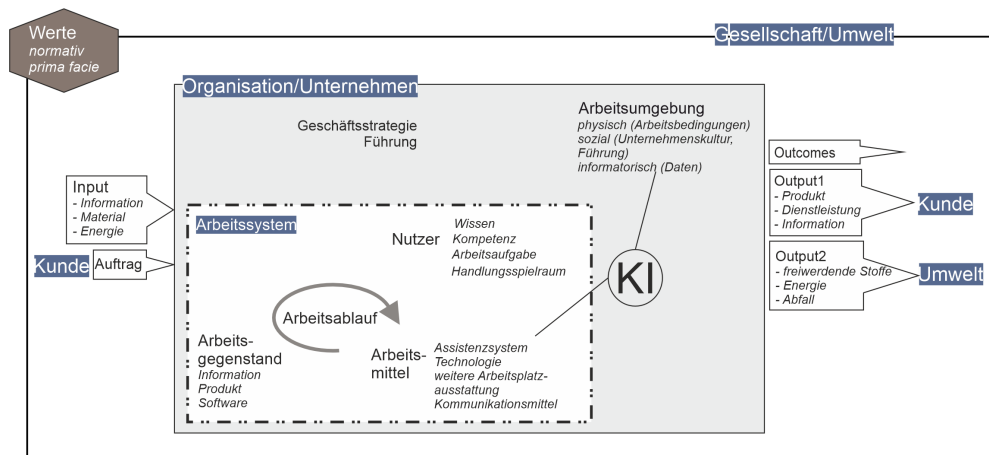


Abbildung 2: *erweiterter Arbeitssystembegriff*

Das entscheidende Arbeitsmittel hinsichtlich einer Bewertung im Arbeitssystem sind die KI-Lösungen. Sie haben quasi eine Doppelfunktion. Einerseits sind sie als Assistenzsystem mit Entscheidungshoheit beim Menschen zu interpretieren, mit dem die Mensch-Technik-Interaktion direkt erfolgt. Andererseits könnten KI-Anwendungen perspektivisch auch autonome Prozesse realisieren, die vermittelt über die Arbeitsumgebung, die Arbeitstätigkeit und die Ausführungsbedingungen beeinflussen, was dann besonders beachtet werden sollte. Es sind bisher wenig empirisch gestützte Erkenntnisse verfügbar, die den konkreten Einfluss der Ausgestaltung von KI auf die Belastungssituation der Beschäftigten zeigen. Hier kommen die Vorteile der Szenarioanalyse hinsichtlich des oben angesprochenen fallspezifischen Vorgehens zum Tragen, indem Vertreter der Arbeitsforschung, der Informatik und der Produktionstechnik mit ihrer Expertise Aufbau und Funktion der KI-Anwendung beschreiben, Use Cases der Interaktion mit der KI-Lösung ableiten, die zu berücksichtigenden Nutzergruppen in Form von Personas charakterisieren und die Einflussfaktoren in ihrer zu erwartenden Ausprägung einschätzen.

Tabelle 1: *Initiale Beschreibung der KI-Anwendung, abgeleitet aus Poretschkin (2021)*

<ul style="list-style-type: none"> • Aufgabenstellung der KI-Anwendung • Eingabedaten und Ausgaben der KI-Anwendung • Monitoring • ML-Modell bzw. Lernalgorithmus • Training und Aktualisierung • Einbettung der KI-Anwendung in ein Gesamtsystem → Interaktion zwischen KI-Anwendung und Gesamtsystem • Aufgaben des Menschen in Bezug auf die KI-Anwendung • funktionale und IT-Sicherheit
--

Erster Vorgehensschritt ist die Beschreibung des KI-Systems, die Formulierung des spezifischen KI-Einsatzes in Form von Use Cases und die Identifizierung potenzieller Personas. Tabelle 1 zeigt die relevanten Merkmale für die Beschreibung der KI-Anwendung (Schritt 1 der Vorgehensweise) auf. Die Use Cases, die mittels bzw. unterstützt durch KI-Anwendungen ausgeführt werden, können anhand vordefinierter Merkmale beschrieben werden, die sich aus dem erweiterten Arbeitssystembegriff ableiten lassen (Tabelle 2). Zur Abbildung unterschiedlicher Nutzeranforderungen werden Nutzergruppen als Personas dargestellt. Damit ist eine Möglichkeit gegeben, sich individueller Arbeitsgestaltung im Entwicklungsprozess anzunähern.

Tabelle 2: *Beschreibungsmerkmal eines Use Cases, abgeleitet aus Paderewski (2020)*

<ul style="list-style-type: none"> • Kurzbeschreibung • Vorbedingungen für die Ausführung • Erwartetes Ergebnis • Reaktionen bei Fehlschlag • Akteure, orientiert an Persona • Auslöser 	<ul style="list-style-type: none"> • Detaillierte Beschreibung des geplanten, idealen Ablaufs (Standardablauf) • Erweiterungen des Standardablaufs • Alternativen für einzelne Arbeitsschritte • Folgehandlungen der Akteure
---	--

Die Persona-Methode erlaubt es, potenzielle zukünftige Nutzer:innen des Assistenzsystems in einer praktikablen und eindeutigen Art zu dokumentieren. Ursprünglich stammt sie aus der Softwareentwicklung und repräsentiert im Gestaltungsprozess der Software die realen Nutzer:innen bzw. eine Nutzergruppe. Eine Persona-Beschreibung ist in der Regel ein Dokument, welches in mehrere Kategorien von Informationen über den fiktiven Nutzer aufgeteilt ist (Abbildung 3). Je nach Einsatzfall wird diese Struktur individuell definiert (Paderewski 2020).

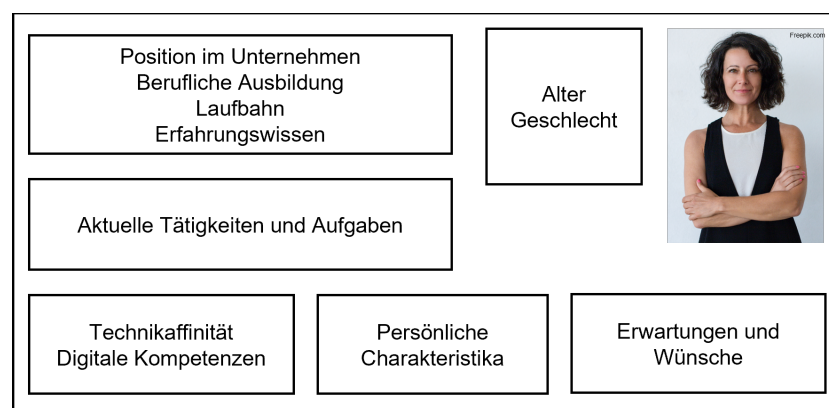


Abbildung 3: *Struktur der Persona-Darstellung, in Anlehnung an Paderewski (2020)*

Für die Identifizierung der Einflussfaktoren in der Szenarioanalyse werden die Einflussbereiche aus der Darstellung des erweiterten Arbeitssystems aufgegriffen und durch KI-relevante Einflussfaktoren untersetzt (exemplarisch in Tabelle 3).

Tabelle 3: *Einflussbereiche und Einflussfaktoren (Gebauer et al. 2022)*

Einflussbereiche	Einflussfaktoren (Beispiele)
Einzelperson	Wissen, Kompetenz, Arbeitsaufgabe, Handlungsspielraum, Affinität zur Technologie, Akzeptanz der Technologie, Erwartungen an die Technologie

Gesellschaft	Szenarien zukünftiger Arbeit, Mitbestimmung, Akzeptanz, Haftungsfragen, Datenschutz
Ausführungsbedingungen	Kompetenzen, IT-Infrastruktur, Datenverfügbarkeit
Arbeitsmittel / Arbeitsgegenstand	Produktions- und Entwicklungskosten, Merkmale der spezifischen Technologie: erweiterte Wahrnehmung und Interaktion, Mensch-Maschine-Koadaption
Unternehmen und Organisation	Grundsätzliche Ausrichtung der Funktionsteilung Mensch - KI-System, Geschäftsstrategie, Führung, Unternehmenskultur, Organisationsform, Qualifizierungsstrategie

Der Bewertungsprozess muss diskursiv für den konkreten Fall erfolgen, da bisher normative Werte für den KI-Einsatz fehlen. Als Erstes sind die Werte auszuhandeln, anhand derer die Lösungsbewertung erfolgen soll. Dazu kann einerseits auf gesicherte arbeitswissenschaftliche Erkenntnisse zurückgegriffen werden, u.a. Usability (EN ISO 14915-2) und Arbeitsgestaltung (DIN EN ISO 6385). Für die ethisch orientierte Bewertung sind andererseits entsprechend des gesellschaftlichen Diskurses Werte auf den Ebenen Individuum, Team, Gesellschaft und Umwelt abzuleiten und zu operationalisieren. Beispielhaft zeigt dies Tabelle 4.

Tabelle 4: Darstellung beispielhafter Werte zur Bewertung der Szenarien (Gebauer et al. 2022)

Konzentrieren Sie sich auf ...	Betroffene Wertebereiche	Beispielwerte
Benutzer	Komfort	Benutzerfreundlichkeit, Körperkomfort, Funktionalität, Praktikabilität
	Vitalität	Gesundheitsfördernder (Langzeit-)Gebrauch, Sicherheit
	Freiheit	Kompetenz, Selbstwirksamkeit, Achtsamkeit, sinnvolles/bewertetes Handeln, intrinsische Motivation, Selbstwertgefühl, Individualität, Privatsphäre, Datenschutz
Team	Wirksamkeit	Effizienz, wirtschaftliche Nachhaltigkeit
	Geselligkeit	Kooperativer Umgang, gegenseitiges Vertrauen, gegenseitiger Respekt
Organisation	Wirksamkeit	Effizienz, wirtschaftliche Nachhaltigkeit
Gesellschaft	Justiz	Offene Gesellschaft, demokratische Beteiligung, Chancen und Beteiligung, Inklusion, Verfügbarkeit "guter" Arbeit
Umgebung	Sicherheit	Ökologische Nachhaltigkeit

An dieser Stelle sind die KI-spezifischen Werte zu integrieren, wie sie u. a. bei Heesen et al. (2020), Wright (2011) und DKE (2020) genannt sind (vgl. Tabelle 5).

Tabelle 5: Von der unabhängigen hochrangigen Expertengruppe für Künstliche Intelligenz identifizierte Werte (DKE 2020)

• Vorrang menschlichen Handelns und menschlicher Aufsicht	• Gesellschaftliches und ökologisches Wohlergehen
• Technische Robustheit und Sicherheit	• Vielfalt, Nichtdiskriminierung, Fairness
• Datenschutz und Datenqualität	• Transparenz und Rechenschaftspflicht

Im Ergebnis der Bewertung erfolgt die Diskussion, insbesondere hinsichtlich ggf. konkurrierender Wertesysteme sowie die Ableitung von Empfehlungen für die weitere Ausgestaltung der Lösung. Ein wichtiges Ergebnis des Diskussionsprozesses ist, die Voraussetzungen und Folgen bestimmter Konstellationen sichtbar und transparent zu machen und alternative Entwicklungspfade zu berücksichtigen.

3. Zusammenfassung und Perspektiven

Die menschenzentrierte Bewertung des KI-Einsatzes ist ein komplexer und sich teilweise wiederholender Prozess. Zu diesem Zweck wird ein begleitendes partizipatives Vorgehensmodell vorgeschlagen, das integraler Bestandteil des Implementationsprozesses ist und in diesem von Beginn an mehrfach mit wachsendem Detaillierungsgrad wiederholt wird. Es unterstützt darüber hinaus notwendige Entwicklungsschritte und Entscheidungen mittels fortlaufender, selbstreflexiver Einschätzungen.

4. Literatur

- Collingridge D (1980): The Social Control of Technology. Pinter, London.
- DKE (2020) ETHIK UND KÜNSTLICHE INTELLIGENZ: Was können technische Normen und Standards leisten? <https://www.dke.de/de/arbeitsfelder/core-safety/whitepaper-ethik-kuenstliche-intelligenz>.
- Gebauer H, Grübler G, Hausmann S, Ott G, Schmauder M (2022) Possible Development Paths of Tactile Internet: Ethical Perspectives, <https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-790863>.
- Grunwald A (2019) Digitalisierung als Prozess. Ethische Herausforderungen inmitten allmählicher Verschiebungen zwischen Mensch, Technik und Gesellschaft. *zfwu*. 2019; 20 (2): 121–45.
- Grunwald A (2002) Technikfolgenabschätzung – eine Einführung. *Nomos*.
- Heesen J et al. (Hrsg.): Ethik-Briefing. Leitfaden für eine verantwortungsvolle Entwicklung und Anwendung von KI-Systemen – Whitepaper aus der Plattform Lernende Systeme, München 2020.
- Kosow H, Gaßner R, Erdmann L (2008). Methoden der Zukunfts- und Szenarioanalyse: Überblick, Bewertung und Auswahlkriterien. Berlin: IZT; 88 S. (WerkstattBericht/IZT, Institut für Zukunftsstudien und Technologiebewertung).
- Kosow H, León CD (2015) Die Szenariotechnik als Methode der Experten- und Stakeholdereinbindung. In: Niederberger M, Wassermann S, Herausgeber. Methoden der Experten- und Stakeholdereinbindung in der sozialwissenschaftlichen Forschung [Internet]. Wiesbaden: Springer Fachmedien Wiesbaden; verfügbar unter: https://doi.org/10.1007/978-3-658-01687-6_11.
- Paderewski J (2020) Entwicklung einer systematischen Vorgehensweise für die Neugestaltung von Arbeitsaufgaben am Beispiel der Einführung eines Assistenzsystems in der Werkstoffprüfung unter Verwendung von Use Cases und Personas, TU Dresden, Diplomarbeit.
- Poretschkin M et al. (2021) Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz [Internet]. [zitiert 11. Januar 2022]. Verfügbar unter: https://www.iais.fraunhofer.de/content/dam/iais/fb/Kuenstliche_intelligenz/ki-pruefkatalog/202107_KI-Pruefkatalog.pdf.
- Wittenbrink, N, Demirci S, Wischmann S (2023) Resiliente und robuste KI-Systeme im praktischen Einsatz. In: iit-Themenband – Resilienz: Leben-Räume-Technik (Wittpahl V (Hrsg.)).
- Wright D (2011) A framework for the ethical impact assessment of information technology. *Ethics Inf Technol* 13: 199–226. doi: 10.1007/s10676-010-9242-6.

Danksagung: Die Darstellung basiert auf Forschungsarbeiten in folgenden geförderten Projekten:

- ETAC Ethical Implications of the Tactile Internet: TU Dresden, Teilaufgabe im Rahmen von CeTI, 2021–2022
- PROKI KI-Demonstrations- und Transferzentrum Umformtechnik, TU Dresden, gefördert vom BMBF 2022–2024
- PAL PerspektiveArbeit Lausitz, Teilprojekt TU Dresden, gefördert vom BMBF 2022–2026



Gesellschaft für Arbeitswissenschaft e.V.

Nachhaltig Arbeiten und Lernen

**Analyse und Gestaltung lernförderlicher
und nachhaltiger Arbeitssysteme
und Arbeits- und Lernprozesse**

69. Kongress der
Gesellschaft für Arbeitswissenschaft e.V.

Gottfried Wilhelm Leibniz Universität Hannover

01. – 03. März 2023

GfA-Press

Bericht zum 69. Arbeitswissenschaftlichen Kongress vom 01. – 03. März 2023

**Fakultät Maschinenbau, Institut für Berufswissenschaften der Metalltechnik (IBM) und
Institut für Fabrikanlagen und Logistik (IFA), Leibniz Universität Hannover**

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Sankt Augustin: GfA-Press, 2023
ISBN 978-3-936804-32-4

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle (s. u.) erhältlich.

Alle Rechte vorbehalten.

© GfA-Press, Sankt Augustin

Schriftleitung: Prof. Dr. Rolf Ellegast

im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Ohne ausdrückliche Genehmigung der Gesellschaft für Arbeitswissenschaft e.V. ist es nicht gestattet:

- den Kongressband oder Teile daraus in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) zu vervielfältigen,
- den Kongressband oder Teile daraus in Print- und/oder Nonprint-Medien (Webseiten, Blog, Social Media) zu verbreiten.

Die Verantwortung für die Inhalte der Beiträge tragen alleine die jeweiligen Verfasser; die GfA haftet nicht für die weitere Verwendung der darin enthaltenen Angaben.

Geschäftsstelle der GfA

Simone John, Tel.: +49 (0)30 1300-13003

Alte Heerstraße 111, D-53757 Sankt Augustin

info@gesellschaft-fuer-arbeitswissenschaft.de · www.gesellschaft-fuer-arbeitswissenschaft.de

Screen design und Umsetzung

© 2023 fröse multimedia, Frank Fröse

office@internetkundenservice.de · www.internetkundenservice.de