

Entwicklung und Validierung des KI-ULTRA Evaluation Toolkit als Selbsteinschätzungswerkzeug für Unternehmen zur Unterstützung des Einsatzes Künstlicher Intelligenz im Arbeitsumfeld

Jan-Paul LEUTERITZ

*Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO,
Nobelstr. 12, D-70569 Stuttgart*

Kurzfassung: Das KI-ULTRA Evaluation Toolkit ist ein für Unternehmen kostenfrei zugängliches Verfahren, um den eigenen Transformationsbedarf hinsichtlich günstiger Rahmenbedingungen für den KI-Einsatz zu messen, und um abgeschlossene KI-Projekte hinsichtlich ihrer Auswirkungen zu bewerten. Hier wird dargestellt, wie das Evaluation Toolkit entwickelt und validiert wurde. In einer Studie mit den teilnehmenden Unternehmenslaboren konnte Evidenz für die Richtigkeit der angenommenen Faktorenstruktur durch exploratorische und konfirmatorische Faktorenanalysen gesammelt werden.

Schlüsselwörter: Künstliche Intelligenz, Transformation, Evaluation, Fragebogen, Rahmenbedingungen

1. Einleitung

Der Einsatz von Anwendungen Künstlicher Intelligenz (KI) im betrieblichen Kontext bietet Unternehmen vielfältige Potenziale. Laut einer Umfrage des Bitkom (2023) sehen 2 Drittel der befragten Unternehmen KI als wichtigste Zukunftstechnologie. Um die Potenziale von KI für sich nutzbar machen zu können, müssen die Unternehmen günstige Rahmenbedingungen schaffen. Im Rahmen des Projekts KI-ULTRA wurde untersucht, wie dies gelingen kann. Das Konsortium hat 29 Organisationen aus Deutschland über knapp 18 Monate bei der Einführung einer KI-Anwendung im Arbeitskontext begleitet. Das Ziel des vom Bundesministerium für Arbeit und Soziales (BMAS) geförderten Projekts bestand darin, herauszufinden, welche konkreten Unterstützungsbedarfe die Unternehmen haben, um die Potenziale von KI bestmöglich für sich im Arbeitsumfeld nutzbar zu machen, sowie mit den Unternehmen geeignete Instrumente zu erarbeiten, um Betrieben aller Größen und Branchen den Einstieg in KI-Technologien zu erleichtern. Diese Instrumente sind der Öffentlichkeit kostenfrei unter <https://www.ergebnisse.ki-ultra.iao.fraunhofer.de> zugänglich. Dazu zählen die beiden KI-ULTRA Leitfäden:

- *Der Leitfaden zu Strategie und Wandel für den KI-Einsatz*
- *Der Leitfaden zur Durchführung von KI-Projekten*

Als Ergänzung zu den Leitfäden wurde im Projekt das aus zwei Modulen bestehende KI-ULTRA Evaluation Toolkit entwickelt. Eine Organisation kann das Modul 1 *Transformationsbedarf* dazu verwenden, festzustellen, inwieweit günstige Rahmenbedingungen für den KI-Einsatz vorliegen und welche Kapitel der Leitfäden für die Weiterentwicklung hilfreich wären. Das Modul 2 *Erfolgsmessung* dient dazu, nach dem Abschluss eines KI-Einführungsprojekts einen Überblick über die Auswirkungen der KI-

Anwendung, insbesondere auf die Mitarbeitenden zu erhalten. Auf der Website können Unternehmen eigenständig Umfragen starten; die Ergebnisse werden automatisiert ausgewertet und unter Berücksichtigung der Datenschutzrichtlinien ausgegeben.

Hier wird beschrieben, wie die beiden Module des Evaluation Toolkit entwickelt wurden und welche Evidenz zur Validität der Skalen bisher vorliegt.

2. Methode: Iterative Entwicklung des Evaluation Toolkit

2.1 Vorgehensweise

Auf Basis einer Literaturanalyse ermittelten wir relevante Zielgrößen. Bei Modul 1 waren dies insbesondere Erfolgsfaktoren der KI-Einführung (z. B. Alsheibani et al. 2019; Chatterjee et al. 2020); bei Modul 2 wurden Kriterien dafür gesammelt, dass sich eine eingeführte KI-Anwendung positiv auf die Mitarbeitenden, die Organisation und die Gesellschaft auswirkt. Über ein fortlaufend aktualisiertes theoretisches Wirkmodell wurde sichergestellt, dass die Leitfäden und das Evaluation Toolkit über die Entwicklungsiterationen kongruent blieben. So war jede Handlungsempfehlung der Leitfäden mit mind. einer messbaren Zielvariable aus dem Evaluation Toolkit verknüpft und vice versa.

Version 1 des Modul 1 des Evaluation Toolkit wurde mit Teilnehmenden aus den beteiligten Organisationen getestet. Anhand der Ergebnisse führten wir für die psychometrischen Konstrukte exploratorische Faktorenanalysen (EFA) durch. Dazu wählten wir eine Hauptachsenanalyse mit Oblimin-Rotation.

Dann wurden mit den teilnehmenden Organisationen jeweils 90-minütige Workshops durchgeführt. Die bis zu 8 Teilnehmenden erarbeiteten hier für beide Module die aus ihrer Sicht nötigen Messgrößen (Variablen, Konstrukte) und Items, indem sie sich fragten, welche Rahmenbedingungen für den erfolgreichen KI-Einsatz aus ihrer Erfahrung nötig sind und welche Erfolgskriterien sie an den KI-Einsatz anlegten. Die Workshops wurden multiperspektivisch besetzt (Mitarbeitende und deren Vertretungen, Projektbeteiligte und Führungskräfte). Anhand der Ergebnisse wurden die bestehenden Entwürfe (Version 1) der beiden Module überarbeitet.

Zum Projektende wurden mit Version 2 des Evaluation Toolkit erneut Befragungsdaten in den teilnehmenden Organisationen gesammelt. Das Modul 2 wurde hier zum ersten Mal getestet, da die meisten Organisationen bei der ersten Erhebung noch keine KI-Anwendung zum Bewerten hatten. Für die Konstrukte des Modul 1 mit als parallel konzipierten Items wurden mit den Daten aus der zweiten Befragungsrunde konfirmatorische Faktorenanalysen (CFA) in *R* gerechnet. Die entsprechend für Modul 2 geplanten exploratorischen Faktorenanalysen konnten aufgrund der kleinen Stichprobe ($N=36$) nicht durchgeführt werden.

2.2 Das Evaluation Toolkit Version 2

Tabellen 1 und 2 stellen die Messgrößen der nun öffentlich verfügbaren Version 2 des Evaluation Toolkit für große Unternehmen (ab 50 Mitarbeitende) dar. Für das Modul 1 gibt es eine eigene Version für kleine Unternehmen mit weniger Dimensionen und teilweise anderen Items. Die Zusammensetzung für eine teilnehmende Person ist von ihrer Rolle im Unternehmen bzw. im Kontext der KI-Einführung abhängig. Dadurch

variiert die Stichprobengröße je nach Konstrukt. Einige Variablen sind als psychometrische Konstrukte angelegt und werden über (idealerweise) parallele Items gemessen. Bei anderen Variablen werden jedoch verschiedene Aspekte abgeprüft, die unabhängig voneinander auftreten können. Beispielsweise erfragen wir zur Messvariable „Interne Unterstützung für Technologieprojekte“, inwieweit verschiedene Akteure (Führung, Betriebsrat, etc.) die Projekte unterstützen. Die Unterstützung eines Akteurs ist unabhängig von der durch einen anderen Akteur. Daher ist hier nicht von parallelen Items auszugehen und eine einfaktorielle Faktorenstruktur der Messvariable stellt kein relevantes Qualitätskriterium dar.

Tabelle 1: Variablen in Version 2 des Modul 1

Variable in Version 2	Parallele Items?	Cronbachs α
Humanistische Organisationskultur	ja	,883
Vertrauens-, Kooperations- und Innovationskultur	ja	,796
Veränderungskultur	ja	,817
Kommunikationskultur (nur in Version 2)	ja	-
Fehlerkultur (nur in Version 2)	ja	-
Datenkultur (nur in Version 2)	ja	,776
Empowerment-orientierte Führung	ja	,731
AI Awareness (Einschätzung von KI-Potenzialen)	nein	-
KI-Mindset (Einstellungen zu KI)	ja	,848
Voraussetzungen für den Projektstart	nein	-
Platzieren von Projektideen	nein	-
Voraussetzungen der Projektdurchführung	nein	-
Alignment von Aktivitäten	nein	-
Voraussetzungen für die Nutzbarmachung der Ergebnisse	nein	-
Interne Unterstützung für Technologieprojekte	nein	-
Organisationsentwicklung (nur in Version 2)	nein	-
Strategieranbindung im Bereich Human Resources	nein	-
Kompetenzen der Benutzenden für den Einsatz der KI am Arbeitsplatz	nein	-
Kompetenzen der Projektbeteiligten für KI-Projekte	nein	-
Kompetenzen für Digitalisierung (nur in Version 2)	nein	-
Persönliche Weiterentwicklung	ja	,743
Veränderungsmotivation (nur in Version 2)	ja	,485
Datenqualität und -quantität	nein	-
Technologieeinsatz (nur in Version 2)	nein	-

Tabelle 2: Variablen in Modul 2 (nur ausgegeben in Version 2)

Variable in Version 2	Parallele Items?
Technologieakzeptanz (Subkonstrukte: Erwartete Leistung, erwarteter Aufwand, Vertrauen, Zufriedenheit, Verhaltensintention)	Ja, in den Subkonstrukten
Veränderung der Tätigkeit hin zu menschlichen Stärken	nein
Erschlossene Qualifikationspotenziale	ja
Autonomie im Umgang mit der KI-Anwendung	ja
Arbeitsplatzbezogene Fertigkeiten und Wissen	ja
Gesellschaftliche Auswirkungen	nein

Arbeitsmarktsituation	nein
Zielerreichung der KI-Lösung	nein
Prozessgestaltung: Transparenz	ja
Prozessgestaltung: Qualität	ja
Veränderung der Arbeitsbedingungen	nein
Aktuelle Arbeitsbedingungen	nein
Ethische Verträglichkeit	nein
Bewertung des Mitwirkungsprozesses	nein

2.3 Stichproben

Am Projekt nahmen 29 Organisationen diverser Branchen und Größen (von unter 10 bis über 10.000 Mitarbeitende) aus Deutschland teil. An der ersten Befragung mit der ersten Version von Modul 1 nahmen $N=115$ Personen teil. In der zweiten Befragungsrunde füllten $N=89$ Personen das Modul 1 aus (davon $N=12$ die Version für kleine Unternehmen) und $N=36$ Personen das Modul 2.

2.4 Gütekriterien der Validierung

Als Voraussetzung der Durchführung einer exploratorischen Faktorenanalyse sollten der Bartlett-Test auf Sphärizität signifikant und der KMO-Index mind. 0,6 betragen. Für Faktoren galt ein Eigenwert von 1 als Untergrenze. Als Voraussetzung der CFA wurden die Daten auf Normalverteilung geprüft und falls notwendig ein robustes Schätzverfahren eingesetzt. Für die CFA akzeptierten wir $TLI > ,95$ und $RMSEA < ,06$ (Hu & Bentler 1999) und das Verhältnis Chi-Quadrat zu Freiheitsgraden (df) bei unter 5 (Schumacker & Lomax 2004).

3. Ergebnisse

3.1 EFA des Modul 1

Die Voraussetzungen für die Durchführung einer EFA bei Modul 1 in der ersten Befragungsrunde waren gegeben für: Organisationskultur, Empowerment-orientierte Führung, KI-Mindset und persönliche Weiterentwicklung. Die EFA der Items zur Organisationskultur lieferte folgende Befunde:

(1) Wir hatten Items mit zwei Arten von Antwortformaten verwendet. Die auf Wertvorstellungen bezogenen Items mit den Antwortankern „ist uns sehr wichtig“ / „ist für uns nicht wichtig“ gliederten sich in eine nachvollziehbare Faktorenstruktur ein. Dagegen zeigte sich bei den auf Grundannahmen bezogenen Items mit den Ankern „ist weitverbreitete Überzeugung“ / „ist eine kaum verbreitete Überzeugung“ ein starker Methodenfaktor, worauf sie aus dem Fragebogen entfernt wurden.

(2) Aufgrund hoher gemeinsamer Faktorladungen wurden die Dimensionen „menschenzentrierte Organisationskultur“ und „chancengleichheitsorientierte Organisationskultur“ zur „humanistischen Organisationskultur“ zusammengefasst. Auch die „Vertrauens-“, die „Kooperations-“ und die „Innovationskultur“ wurden zu einem einzigen Faktor verschmolzen.

Beim KI-Mindset wurden mehrere Faktoren identifiziert und in Version 2 nur die Items desjenigen Faktors übernommen, welche die generelle Einstellung gegenüber

KI adressieren, unabhängig von erwarteten Auswirkungen der KI oder der Mitwirkungsintention. Bei allen anderen Konstrukten wurde eine eindimensionale Struktur bestätigt.

3.2 Ergänzungen anhand der Workshops

Anhand der Workshopergebnisse wurden folgende Variablen zu Modul 1 hinzugefügt: (1) Kommunikationskultur; (2) Fehlerkultur; (3) Datenkultur; (4) Organisationsentwicklung; (5) Veränderungsmotivation; (6) Technologieeinsatz, (7) Kompetenzen für Digitalisierung.

3.3 CFA des Modul 1

Die Stichprobengröße erlaubte es nicht, das gesamte Messmodell aller psychometrischen Variablen in einer einzigen CFA zu testen. Daher wurden einzelne CFAs für die jeweils inhaltlich zusammenhängenden Konstrukte (z. B. alle Dimensionen der Organisationskultur berechnet). Die Ergebnisse der CFAs sind in Tabelle 3 zusammengefasst. Bei der Organisationskultur zeigte das Modell der Variablen (1) Humanistische Organisationskultur, (2) Veränderungs- Kooperations- und Innovationskultur, (3) Veränderungskultur, und (4) Datenkultur einen sehr guten Fit. Dabei wurden 8 Items von der Analyse ausgeschlossen. Wurden die neuen Konstrukte Kommunikationskultur und Fehlerkultur mit einbezogen, so zeigte sich ein Fit unterhalb der definierten Grenzwerte, da die Items dieser Dimensionen zu stark mit jenen aus den ersten beiden Dimensionen korrelierten. Die Messmodelle für die Empowerment-orientierte Führung und das KI-Mindset bestätigten sich als eindimensional. Ein letztes Modell schloss (1) Persönliche Weiterentwicklung und (2) Veränderungsmotivation ein. Auch hier wurden die Qualitätskriterien erfüllt.

Tabelle 3: Fit-Indizes der CFAs zu den psychometrischen Konstrukten in Modul 1

Variablen	Chi-Quadrat(df), <i>p</i>	Chi-Quadrat / Freiheitsgrade	TLI	Robust RMSEA
Organisationskultur	35,150(36), <i>p</i> = ,509	,976	1,000	,029
Empowerment-orientierte Führung	1,292(2), <i>p</i> = ,524	,646	1,000	,000
KI-Mindset	0,844 (2), <i>p</i> = ,656	,422	1,000	,000
Weiterentwicklung & Veränderungsmotivation	12,353 (12), <i>p</i> = ,418	1,029	,990	,023

4. Diskussion

Bis auf Cronbachs α der Veränderungsmotivation erfüllen alle Teststatistiken die Erwartungen an gute Modell-Fits und akzeptable Reliabilitäten. Es wurden einige Items ausgeschlossen, allerdings ist dies wünschenswert, da es den Fragebogen verkürzt und damit besser für die Anwendung in Unternehmen geeignet macht. Die Items zu Kommunikationskultur und Fehlerkultur wären aus messtheoretischer Sicht verzichtbar, können aber für Unternehmen möglicherweise Handlungsmöglichkeiten aufzeigen. Folgende Limitationen dieser Studie sind zu berücksichtigen.

Die bisher verfügbaren Stichproben sind zu klein, um ein abschließendes Urteil über die Validität der entwickelten Fragebögen abzugeben. Insbesondere waren am Ende des Projekts KI-ULTRA noch nicht alle teilnehmenden Unternehmen mit ihrem KI-Projekt weit genug fortgeschritten, um eine Erfolgsbewertung vorzunehmen. Wenn über die Online-Plattform eine ausreichende Menge an Daten erfasst werden kann, so sollten CFAs des gesamten psychometrischen Messmodells durchgeführt werden. Dennoch weisen die vorhandenen Daten darauf hin, dass die psychometrischen Messmodelle funktionieren.

Des Weiteren ist hervorzuheben, dass die Konstrukte nicht anhand von externen Kriterien oder anderen Fragebögen validiert werden konnten. Im Projektverlauf war es wichtig, auf die Bedarfe der teilnehmenden Unternehmen einzugehen und eine derart aufwendige Erfassung von Validierungskriterien war in diesem Kontext nicht möglich. Insbesondere die teilnehmenden KMU haben im Rahmen einer KI-Einführung erhebliche Aufwände zu bewältigen und daher nur geringes Interesse an der Erhebung von Daten, die für sie nicht direkt handlungsrelevant sind. Es ist anzustreben, die Instrumente mit externen Kriterien zu validieren.

Auf Basis der erhobenen Daten kann der Fragebogen der Öffentlichkeit zur Verfügung gestellt werden.

5. Literatur

- Alsheibani S, Cheung Y, Messom C (2019) Factors Inhibiting the Adoption of Artificial Intelligence at organizational-level: A Preliminary Investigation. Twenty-fifth Americas Conference on Information Systems, Cancun.
- Bitkom (2023). Presseinformation: Deutsche Wirtschaft drückt bei Künstlicher Intelligenz aufs Tempo. Accessed Jan 12, 2024. <https://www.bitkom.org/Presse/Presseinformation/Deutsche-Wirtschaft-drueckt-bei-Kuenstlicher-Intelligenz-aufs-Tempo>
- Chatterjee S, Tamilmani K, Rana NP, Dwivedi YK (2020) Employees' Acceptance of AI Integrated CRM System: Development of a Conceptual Model. International Working Conference on Transfer and Diffusion of IT (TDIT), Dec 2020, Tiruchirappalli, India. 679-687, DOI: 10.1007/978-3-030-64861-9_59.
- Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1):1-55. <https://doi.org/10.1080/10705519909540118>.
- Schumacker RE, Lomax RG (2004) A beginner's guide to structural equation modeling (2nd ed.). Mahwah: Lawrence Erlbaum Associates.

Danksagung: Das KI-ULTRA Projektteam bedankt sich herzlich bei den teilnehmenden Unternehmenslaboren für die wertvolle Zusammenarbeit.



Gesellschaft für Arbeitswissenschaft e.V.

Arbeitswissenschaft in-the-loop

**Mensch-Technologie-Integration
und ihre Auswirkung auf Mensch,
Arbeit und Arbeitsgestaltung**

70. Kongress der
Gesellschaft für Arbeitswissenschaft e.V.

Institut für Arbeitswissenschaft und
Technologiemanagement IAT
Universität Stuttgart

In Zusammenarbeit mit dem Fraunhofer-Institut für
Arbeitswirtschaft und Organisation IAO

06. – 08. März 2024

GfA-Press

Bericht zum 70. Arbeitswissenschaftlichen Kongress vom 06. – 08. März 2024

Institut für Arbeitswissenschaft und Technologiemanagement (IAT), Universität Stuttgart

In Zusammenarbeit mit: Fraunhofer-Institut für Arbeitswirtschaft und Organisation (IAO), Stuttgart

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.

Sankt Augustin: GfA-Press, 2024

ISBN 978-3-936804-34-8

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

© 2024 fröse multimedia, Frank Fröse,

office@internetkundenservice.de, www.internetkundenservice.de