Die dimensionale Struktur der KI-Kompetenz

Was Unternehmen wirklich messen, wenn sie „KI-Kompetenz" messen – eine psychologisch fundierte Landkarte der wichtigsten Verfahren im internationalen und deutschsprachigen Raum

Dr. Walter Lieberei

22.07.2026

1. Einleitung: KI-Kompetenz als Überlebensfrage – und als ungelöstes Definitionsproblem

Kaum eine Fähigkeit wird derzeit so häufig als „Schlüsselkompetenz des 21. Jahrhunderts" bezeichnet wie der kompetente Umgang mit Künstlicher Intelligenz (KI). Die Gründe dafür sind handfest: Generative KI-Systeme – also Systeme, die auf Basis großer Sprach- und Bildmodelle eigenständig Texte, Analysen, Programmcode oder Bilder erzeugen – haben innerhalb weniger Jahre Einzug in nahezu alle Wertschöpfungsprozesse gehalten. Unternehmen, deren Belegschaften diese Werkzeuge produktiv, kritisch und verantwortungsvoll nutzen können, realisieren messbare Produktivitäts- und Qualitätsvorteile. Unternehmen, denen diese Fähigkeit fehlt, riskieren dagegen nicht nur Effizienznachteile, sondern auch handfeste Schäden: fehlerhafte, ungeprüft übernommene KI-Ergebnisse, Datenschutzverstöße, Reputationsrisiken und strategische Fehlentscheidungen. KI-Kompetenz ist damit längst keine „Nice-to-have"-Qualifikation mehr, sondern ein Faktor der organisationalen Überlebensfähigkeit.

Der Gesetzgeber hat diese Entwicklung nachvollzogen. Mit Artikel 4 der europäischen KI-Verordnung (englisch: Artificial Intelligence Act, kurz AI Act; Verordnung (EU) 2024/1689) gilt seit dem 2. Februar 2025 für Anbieter und Betreiber von KI-Systemen die Pflicht, „nach besten Kräften sicherzustellen, dass ihr Personal und andere Personen, die in ihrem Auftrag mit dem Betrieb und der Nutzung von KI-Systemen befasst sind, über ein ausreichendes Maß an KI-Kompetenz verfügen". Damit ist KI-Kompetenz – im Verordnungstext als „AI Literacy" bezeichnet – erstmals eine regulatorische Anforderung, die faktisch jedes Unternehmen in der Europäischen Union betrifft, das KI-Systeme einsetzt. Bemerkenswert ist allerdings: Die Verordnung schreibt zwar die Pflicht fest, definiert aber weder verbindlich, welche Teilkompetenzen KI-Kompetenz genau umfasst, noch, wie ein „ausreichendes Maß" festzustellen und nachzuweisen wäre.

Genau an dieser Stelle beginnt das Problem, dem sich dieser Artikel widmet. Das Feld der Künstlichen Intelligenz entwickelt sich in einem Tempo, das klassische Kompetenzdefinitionen systematisch überholt: Was vor achtzehn Monaten als fortgeschrittene Fertigkeit galt – etwa das kunstvolle Formulieren von Eingabeaufforderungen (sogenanntes Prompting bzw. Prompt Engineering) –, wird durch leistungsfähigere, fehlertolerantere Modelle teilweise schon wieder entwertet. Gleichzeitig entstehen laufend neue Anforderungen, etwa im Umgang mit autonom handelnden KI-Agenten oder mit KI-generierten Falschinhalten. Für Unternehmen und Organisationen bedeutet das ein echtes Dilemma: Sie sollen KI-Kompetenz aufbauen, sicherstellen und nachweisen – während die Wissenschaft noch darüber verhandelt, was KI-Kompetenz eigentlich genau ist und wie man sie zuverlässig misst.

Die gute Nachricht lautet: Die psychologische und psychometrische Forschung hat in den Jahren 2023 bis 2026 erhebliche Fortschritte gemacht. Es liegen inzwischen Dutzende wissenschaftlich konstruierte Messverfahren vor, und über deren Struktur zeichnet sich ein erkennbarer – wenn auch noch kein endgültiger – Konsens ab. Eine systematische Übersichtsarbeit (Review) identifizierte bereits 2024 auf Basis von 22 Validierungsstudien insgesamt 16 verschiedene psychometrisch geprüfte KI-Kompetenz-Ska-

len, von denen allerdings erst drei leistungsbasiert, also als objektive Tests, angelegt waren (Lintner, 2024). Seither ist insbesondere die Zahl objektiver Verfahren deutlich gewachsen. Dieser Artikel ordnet das Feld: Er erläutert zunächst, wie sich KI-Kompetenz aus psychologischer Sicht strukturieren lässt, stellt dann die wichtigsten Selbst- und Fremdeinschätzungsverfahren im internationalen und im deutschsprachigen Raum vor, zieht eine Bilanz zur derzeit verbreitetsten dimensionalen Struktur und leitet daraus konkrete Empfehlungen für die kommenden zwei bis drei Jahre ab.

2. KI-Kompetenz aus psychologischer Sicht: Wissen und Fertigkeiten

Wer über die Messung von KI-Kompetenz sprechen will, braucht zunächst begriffliche Klarheit. In der pädagogisch-psychologischen Forschung wird Kompetenz überwiegend als kontextspezifische, erlernbare kognitive Leistungsdisposition verstanden – also als die grundsätzliche Befähigung einer Person, Anforderungen in einem bestimmten Inhaltsbereich erfolgreich zu bewältigen. Auf dieser Definitionstradition bauen auch die aktuellen KI-Kompetenzmodelle auf, etwa das 2026 publizierte AIEDTEC-Rahmenmodell (AIEDTEC steht für „Competence to Use Artificial Intelligence and Digital Technology in Educational Processes", also die Kompetenz zur Nutzung von KI und digitaler Technologie in Bildungsprozessen) einer Frankfurter Forschungsgruppe um Andreas Frey (Frey et al., 2026).

Zerlegt man das Konstrukt KI-Kompetenz psychologisch, ergeben sich zwei grundlegende Bestandteile, die in praktisch allen seriösen Kompetenzmodellen wiederkehren.

Der erste Bestandteil ist das **Wissen über KI**. Die Psychologie unterscheidet hier traditionell zwei Wissensarten. **Deklaratives Wissen** ist das „Wissen, dass": Eine Person weiß beispielsweise, dass große Sprachmodelle auf Wahrscheinlichkeitsberechnungen über Wortfolgen beruhen, dass sie aus Trainingsdaten lernen, dass sie „halluzinieren" können (also überzeugend klingende, aber sachlich falsche Inhalte erzeugen) und dass die europäische KI-Verordnung bestimmte Anwendungen verbietet. **Prozedurales Wissen** ist dagegen das „Wissen, wie": das Wissen um Handlungsabläufe und Vorgehensweisen, etwa wie man eine Aufgabe sinnvoll in Teilschritte für ein KI-System zerlegt, wie man ein Ergebnis systematisch gegenprüft oder wie man ein KI-Werkzeug in einen bestehenden Arbeitsprozess einbettet. Beide Wissensarten zusammen bilden das theoretische Fundament der KI-Kompetenz – sie lassen sich lehren, in Schulungen vermitteln und mit Wissensfragen prüfen.

Der zweite Bestandteil sind die **Fertigkeiten** (englisch: Skills). Fertigkeiten bezeichnen die konkrete Anwendung und Umsetzung des Wissens in der Praxis – also das tatsächliche Können im Vollzug. Eine Person kann theoretisch wissen, wie gutes Prompting funktioniert, und dennoch in der realen Arbeitssituation unbrauchbare Eingaben formulieren. Umgekehrt entwickeln manche Vielnutzer beachtliche praktische Routinen, ohne die dahinterliegenden Konzepte erklären zu können. Psychologisch gesprochen: Wissen ist eine notwendige, aber keine hinreichende Bedingung für kompetentes Handeln. Erst wenn deklaratives und prozedurales Wissen in flüssiges, situationsangemessenes Handeln übersetzt werden, sprechen wir von Fertigkeiten – und erst das Zusammenspiel von Wissen und Fertigkeiten (ergänzt um motivationale und einstellungsbezogene Facetten, die hier nicht im Zentrum stehen) ergibt Kompetenz.

Diese Zweiteilung ist keineswegs akademische Haarspalterei, sondern hat unmittelbare Konsequenzen für die Personalarbeit. Wissensdefizite lassen sich vergleichsweise schnell durch Schulungen beheben; Fertigkeitsdefizite erfordern dagegen Übung, Feedback und begleitete Praxisanwendung. Ein Unternehmen, das nur Wissen vermittelt (etwa durch Pflicht-E-Learnings), aber keine Fertigkeiten aufbaut, erfüllt möglicherweise formal seine Schulungspflicht – erzeugt aber keine handlungswirksame Kompetenz. Und umgekehrt gilt: Ein Unternehmen, das nicht zwischen beiden Bestandteilen unterscheidet, kann auch nicht sinnvoll diagnostizieren, wo seine Belegschaft tatsächlich steht.

Moderne Kompetenzmodelle verfeinern diese Grundstruktur zusätzlich, indem sie Wissensinhalte mit kognitiven Anforderungsniveaus kreuzen. Das erwähnte AIEDTEC-Modell etwa verbindet vier Wissensbereiche (in Anlehnung an das TPACK-Modell, das „Technological Pedagogical Content Knowledge", also technologisches, pädagogisches und inhaltliches Wissen integriert) mit den sechs kognitiven Prozessen der revidierten Bloom-Taxonomie – Erinnern, Verstehen, Anwenden, Analysieren, Bewerten und Erschaffen – und beschreibt darüber hinaus einen vierphasigen Entwicklungsverlauf vom Grundlagenwissen bis zur eigenständigen Gestaltung KI-gestützter Prozesse (Frey et al., 2026). Diese Matrixlogik – Wissensart mal kognitive Operation – gilt heute als tragfähigste Grundlage für die systematische Konstruktion von Kompetenztests, weil sie sicherstellt, dass ein Test nicht nur Faktenwissen auf niedrigem Niveau abfragt, sondern auch anspruchsvolle Analyse-, Bewertungs- und Gestaltungsleistungen erfasst.

3. Wie lässt sich KI-Kompetenz messen? Selbsteinschätzung versus Fremdeinschätzung

Steht fest, *was* gemessen werden soll, stellt sich die zweite Grundsatzfrage: *Wie* wird gemessen? Die Psychometrie – die wissenschaftliche Disziplin der psychologischen Messung – unterscheidet bei der Erfassung von KI-Wissen und KI-Fertigkeiten zwei grundlegend verschiedene Messzugänge, deren Unterschied für die Praxis gar nicht überschätzt werden kann.

Der erste Zugang ist die **Selbsteinschätzung** (englisch: Self-Report oder Self-Assessment). Hier beurteilen Personen ihre eigene Kompetenz, typischerweise anhand sogenannter Likert-Skalen. Eine Likert-Skala (benannt nach dem amerikanischen Sozialpsychologen Rensis Likert) ist ein Antwortformat, bei dem Aussagen wie „Ich kann die Ergebnisse eines KI-Systems kritisch hinterfragen" auf einer mehrstufigen Zustimmungsskala bewertet werden, etwa von 1 („trifft überhaupt nicht zu") bis 5 („trifft voll und ganz zu"). Selbsteinschätzungsskalen sind ökonomisch, schnell durchführbar und für große Stichproben geeignet. Ihr grundsätzliches Problem: Sie messen streng genommen nicht die Kompetenz selbst, sondern die *wahrgenommene* Kompetenz – also ein Selbstbild, das von Selbstwirksamkeitserwartungen, Persönlichkeitsmerkmalen, sozialer Erwünschtheit und systematischen Urteilsverzerrungen überlagert wird. Besonders bekannt ist der Dunning-Kruger-Effekt, wonach gerade Personen mit geringer tatsächlicher Kompetenz ihre Fähigkeiten überschätzen, weil ihnen das Wissen fehlt, die eigenen Defizite zu erkennen. Selbsteinschätzungen liefern deshalb bestenfalls vage Näherungswerte für tatsächliches Wissen und Können.

Der zweite Zugang ist die **Fremdeinschätzung im weiteren Sinne**, präziser: die *objektive Leistungsmessung*. Hier wird Kompetenz nicht erfragt, sondern demonstriert. Die gängigsten Formate sind klassische Test-Items mit richtigen und falschen Antworten (meist im Multiple-Choice-Format, kurz MC, also Auswahlfragen mit vorgegebenen Antwortoptionen), Situational-Judgement-Items (situative Urteilsaufgaben, kurz SJT für „Situational Judgement Test", bei denen realistische Arbeitssituationen geschildert und Handlungsoptionen bewertet werden müssen) sowie performanzbasierte Aufgaben, bei denen Personen tatsächliche KI-Interaktionen durchführen und deren Ergebnisqualität bewertet wird. Objektive Verfahren sind aufwendiger zu konstruieren und durchzuführen, gelten aber als deutlich zuverlässiger, weil sie tatsächliches Wissen und tatsächliche Urteilsfähigkeit erfassen und gegen Selbstüberschätzung immun sind.

Wie groß die Kluft zwischen beiden Messzugängen ist, gehört zu den wichtigsten – und für die Praxis unbequemsten – Befunden der jüngsten Forschung. Studien, die beide Zugänge direkt vergleichen, berichten durchweg nur geringe Zusammenhänge zwischen selbsteingeschätzter und objektiv gemessener KI-Kompetenz. Die Entwickler des Generative AI Literacy Assessment Test (GLAT) konnten zudem zeigen, dass objektive Testwerte die reale Leistung in KI-gestützten Aufgaben besser vorhersagen als Selbsteinschätzungen (Jin et al., 2025). Wer also Personalentscheidungen, Zertifizierungen oder den

Nachweis nach Artikel 4 der KI-Verordnung allein auf Selbstauskünfte stützt, misst zu einem erheblichen Teil Selbstvertrauen statt Kompetenz.

Die folgende Tabelle fasst die beiden Messzugänge zusammen.

Tabelle 1: Messzugänge zur Erfassung von KI-Kompetenz im Vergleich

Merkmal	Selbsteinschätzung	Fremdeinschätzung / objektive Messung
Typisches Format	Likert-Skalen („Ich kann …")	Multiple-Choice-Items, Situational-Judgement-Items, Performanzaufgaben
Gemessenes Konstrukt	Wahrgenommene Kompetenz, Selbstwirksamkeit, Nutzungssicherheit	Demonstriertes Wissen, Urteils- und Handlungsfähigkeit
Durchführungsaufwand	Gering (oft unter 10 Minuten)	Mittel bis hoch (Testkonstruktion, Durchführung, Auswertung)
Typische Verzerrungen	Selbstüberschätzung (Dunning-Kruger-Effekt), soziale Erwünschtheit, Referenzgruppeneffekte	Testangst, Rateverhalten (durch Testdesign kontrollierbar)
Zusammenhang mit realer Leistung	Gering bis moderat	Deutlich höher; beste Evidenz bei performanznahen Verfahren
Geeignet für	Monitoring, Bedarfserhebung, Evaluation von Trainings, Forschung	Kompetenzdiagnostik, Personalauswahl, Zertifizierung, Compliance-Nachweis
Nicht geeignet für	Individualdiagnostik, Auswahl- und Zertifizierungsentscheidungen	– (Grenzen eher bei Aufwand und Aktualität der Items)

Für die Unternehmenspraxis folgt daraus keine pauschale Abwertung der Selbsteinschätzung – sie hat ihren legitimen Platz, etwa in der Bedarfsanalyse, im Kulturmonitoring oder zur Erfassung motivationaler Facetten wie KI-Angst oder Veränderungsbereitschaft. Wo aber belastbare Aussagen über tatsächliches Wissen und tatsächliche Fertigkeiten gebraucht werden, führt an objektiven oder simulationsbasierten Verfahren kein Weg vorbei.

4. Die wichtigsten Verfahren im internationalen Umfeld

Die internationale Instrumentenlandschaft hat sich seit 2023 rasant entwickelt. Im Folgenden werden die bekanntesten und wissenschaftlich am besten dokumentierten Verfahren vorgestellt – getrennt nach Selbsteinschätzungs- und Fremdeinschätzungsverfahren – jeweils mit ihrer dimensionalen Struktur, also der Frage, in welche Teilkompetenzen das jeweilige Verfahren KI-Kompetenz zerlegt.

4.1 Selbsteinschätzungsverfahren

Das international wohl meistzitierte Selbsteinschätzungsverfahren ist die **MAILS – Meta AI Literacy Scale** (Carolus et al., 2023), entwickelt an der Universität Würzburg, aber englischsprachig publiziert und inzwischen in zahlreichen Ländern eingesetzt und weitervalidiert. Die MAILS basiert konzeptionell auf dem einflussreichen Kompetenzkatalog von Ng und Kollegen (Ng et al., 2021) und erweitert ihn um psychologische Meta-Kompetenzen. Die Vollversion umfasst 34 Items und misst KI-Kompetenz über vier übergeordnete Bereiche: erstens „AI Literacy" im engeren Sinne mit den Facetten KI verwenden und anwenden, KI verstehen, KI erkennen und KI-Ethik; zweitens „KI-Selbstwirksamkeit" (die Überzeugung, KI-bezogene Anforderungen bewältigen zu können) mit den Facetten Problemlösen und Lernen; drittens „KI-Selbstkompetenz" mit den psychologischen Meta-Facetten Emotionsregulation und Persistenz im Umgang mit KI; viertens die Facette „KI-Gestaltung" (Programmieren und Entwickeln von KI). Bemerkenswert an der MAILS ist gerade diese Erweiterung um nicht-kognitive Facetten – sie trägt der Einsicht

Rechnung, dass kompetentes KI-Handeln auch emotionale Stabilität und Lernbereitschaft im Umgang mit einer sich schnell wandelnden Technologie voraussetzt.

Das zweite international etablierte Selbsteinschätzungsverfahren ist die **SNAIL – Scale for the Assessment of Non-Experts' AI Literacy** (Laupichler et al., 2023), entwickelt an der Universität Bonn. Die SNAIL richtet sich ausdrücklich an Nicht-Experten, also Personen ohne Informatik-Hintergrund, und umfasst in der validierten Fassung 31 Items auf drei Dimensionen: „Technisches Verständnis" (Technical Understanding), „Kritische Bewertung" (Critical Appraisal) und „Praktische Anwendung" (Practical Application). Die dreifaktorielle Struktur wurde mittels explorativer Faktorenanalyse gewonnen – einem statistischen Verfahren, das aus den Antwortmustern vieler Personen die zugrundeliegenden Dimensionen extrahiert – und bildet die klassische Trias aus Wissen, Bewertung und Anwendung ab.

Neben diesen beiden Referenzverfahren existiert eine inzwischen kaum mehr überschaubare Zahl weiterer Selbstauskunftsskalen. Das erwähnte systematische Review (Lintner, 2024) wertete 22 Validierungsstudien zu 16 verschiedenen Skalen aus und kam zu einem ernüchternden Zwischenfazit: Die meisten Skalen adressieren Studierende oder die Allgemeinbevölkerung, die Skalen zeigen zwar überwiegend gute Strukturvalidität und interne Konsistenz, doch Belege für kulturübergreifende Gültigkeit und Messinvarianz (also die Sicherstellung, dass ein Test in verschiedenen Gruppen dasselbe misst) fehlen fast vollständig – und 13 der 16 Skalen beruhen auf Selbstauskunft. Seit 2024 sind zudem zahlreiche berufsgruppenspezifische Selbsteinschätzungsskalen hinzugekommen, etwa für Pflegekräfte, Lehrkräfte und Hochschulpersonal, die überwiegend vier bis sechs Dimensionen unterscheiden – typischerweise Grundlagenwissen, Anwendung, kritische Bewertung, ethisch-rechtliche Verantwortung sowie zunehmend eine eigene Dimension für generative KI und für kontinuierliche Kompetenzaktualisierung.

Tabelle 2: Bekannte internationale Selbsteinschätzungsverfahren und ihre dimensionale Struktur

Verfahren	Quelle	Items	Dimensionale Struktur
MAILS (Meta AI Literacy Scale)	Carolus et al. (2023)	34	AI Literacy (Verwenden & Anwenden, Verstehen, Erkennen, Ethik), KI-Selbstwirksamkeit (Problemlösen, Lernen), KI-Selbstkompetenz (Emotionsregulation, Persistenz), KI-Gestaltung
SNAIL (Scale for the Assessment of Non-Experts' AI Literacy)	Laupichler et al. (2023)	31	Technisches Verständnis, Kritische Bewertung, Praktische Anwendung
Diverse GenAI- und berufsspezifische Skalen (2024–2026)	Überblick in Lintner (2024)	meist 15–35	überwiegend 3–6 Faktoren; wiederkehrend: Wissen/Verstehen, Anwenden, Bewerten, Ethik; zunehmend ergänzt um GenAI-Nutzung und Kompetenzaktualisierung

4.2 Fremdeinschätzungs- und Leistungstestverfahren

Der Referenzpunkt unter den objektiven Verfahren ist der **AI Literacy Test** von Hornberger, Bewersdorff und Nerdel (Hornberger et al., 2023), entwickelt an der Technischen Universität München (TUM). Der Test umfasst 30 Multiple-Choice-Items (plus ein Ankeritem) und deckt entlang etablierter KI-Literacy-Rahmenmodelle – insbesondere der 17 Kompetenzen von Long und Magerko (2020) – ein breites Themenspektrum ab: Grundkonzepte der KI, maschinelles Lernen, praktische Anwendungen, gesellschaftliche Auswirkungen und Ethik. Psychometrisch bemerkenswert: Obwohl der Test inhaltlich mehrere Kompetenzbereiche abdeckt, erwies er sich empirisch als im Wesentlichen *eindimensional* – die Leistung lässt sich sparsam durch einen einzigen übergeordneten Fähigkeitsfaktor beschreiben. Dieser Befund zieht sich wie ein roter Faden durch die objektive Testforschung und wird uns in der Bilanz (Kapitel 6) wieder begegnen.

Aus demselben Forschungsprogramm stammt der **AILIT-S**, eine 2025 publizierte Kurzform mit nur 10 Items, die in einer internationalen Stichprobe von 1.465 Studierenden aus Deutschland, Großbritannien und den USA validiert wurde (Hornberger et al., 2025). Die Autoren begrenzen den Einsatzzweck ausdrücklich auf Gruppenvergleiche, Monitoring und Forschung – für individualdiagnostische Entscheidungen ist eine derart kurze Skala nicht ausreichend messgenau. Diese Selbstbeschränkung ist vorbildlich und zugleich eine wichtige Warnung an die Praxis: Nicht jedes wissenschaftlich validierte Instrument ist für jeden Zweck geeignet.

Für den Bereich der generativen KI ist der **GLAT – Generative AI Literacy Assessment Test** (Jin et al., 2025) das derzeit prominenteste objektive Verfahren. Der GLAT umfasst 20 szenariobasierte Multiple-Choice-Items und folgt einer vierdimensionalen Inhaltsstruktur: „Kennen und Verstehen" (Know & Understand), „Nutzen und Anwenden" (Use & Apply), „Bewerten und Gestalten" (Evaluate & Create) sowie „Ethik" (Ethics). Diese vier Dimensionen gehen direkt auf das Rahmenmodell von Ng et al. (2021) zurück, das sich damit als das international einflussreichste Strukturmodell erweist. Der GLAT wurde sowohl klassisch testtheoretisch als auch mit Verfahren der Item-Response-Theorie (IRT; einer modernen testtheoretischen Methodenfamilie, die Itemschwierigkeit und Personenfähigkeit auf einer gemeinsamen Skala modelliert) geprüft. Sein wichtigster Validitätsbeleg wurde bereits erwähnt: GLAT-Werte sagen reale Leistungen in KI-gestützten Aufgaben besser vorher als Selbsteinschätzungen.

Das derzeit umfassendste objektive Verfahren ist die **AICOS – AI Competency Objective Scale** (Markus et al., 2025), ebenfalls aus dem Würzburger Umfeld. Die AICOS operationalisiert KI-Kompetenz über sechs Subkompetenzen: KI verstehen, KI anwenden, KI erkennen, KI erzeugen bzw. gestalten, KI-Ethik sowie – als eigenständiges Modul – generative KI. Der Volltest umfasst 51 Multiple-Choice-Items, eine Kurzform 18 Items. Die AICOS verbindet damit die klassische KI-Literacy-Struktur mit einer expliziten GenAI-Dimension und ist ausdrücklich modular angelegt. Offen ist nach Einschätzung der Autoren selbst, ob generative KI langfristig eine eigenständige Dimension bleiben oder als Querschnittsthema in alle anderen Facetten integriert werden sollte – eine Frage, die die gesamte Testlandschaft derzeit beschäftigt.

Einen thematisch breiten Zugang für die erwachsene Allgemeinbevölkerung bietet die **SAIL4ALL – Scale of Artificial Intelligence Literacy for All** (Soto-Sanfiel et al., 2025), ein wissensbasierter Test mit Wahr/Falsch-Format, der in vier Themenblöcke gegliedert ist: „Was ist KI?", „Was kann KI?", „Wie funktioniert KI?" und „Wie sollte KI genutzt werden?". Psychometrisch interessant ist, dass die ersten beiden Themenblöcke empirisch jeweils in zwei Subfaktoren zerfallen, sodass insgesamt fünf Auswertungsscores resultieren. Die SAIL4ALL zeigt damit exemplarisch, dass thematische Gliederung und empirische Faktorenstruktur zwei verschiedene Dinge sind – ein Test kann inhaltlich vier Themen abdecken, aber statistisch eine andere Binnenstruktur aufweisen.

Einen konzeptionell eigenständigen Weg geht schließlich die Arbeitsgruppe um Ning Li von der Tsinghua-Universität mit ihrem **A-Faktor-Ansatz** (Li et al., 2025). In Analogie zum Generalfaktor der Intelligenz (dem sogenannten g-Faktor nach Charles Spearman: der empirische Befund, dass Leistungen in ganz unterschiedlichen Denkaufgaben positiv korrelieren und sich durch einen gemeinsamen Faktor erklären lassen) postulieren und belegen die Autoren einen „A-Factor" – einen allgemeinen Faktor der KI-Interaktionskompetenz. In drei Studien mit insgesamt 517 Teilnehmenden bearbeiteten Personen simulierte und reale Aufgaben der Zusammenarbeit mit generativer KI; die Leistungen wurden durch menschliche Bewerter und durch GPT-4 beurteilt. Ein dominanter erster Faktor erklärte rund 44 Prozent der Leistungsvarianz über alle Aufgaben hinweg. Darunter identifizierten die Autoren vier Erstordnungsdimensionen: Kommunikationseffektivität mit KI, kreative Ideengenerierung, Bewertung KI-generierter Inhalte sowie schrittweise Kollaboration bzw. Problemzerlegung. Das Endprodukt ist eine performanzbasierte 18-Aufgaben-Batterie mit hierarchischer Struktur – vier Dimensionen unter einem Generalfaktor – und sehr gutem Modellfit in der konfirmatorischen Faktorenanalyse. Der A-Faktor-Ansatz ist deshalb be-

deutsam, weil er die KI-Kompetenzmessung methodisch an die jahrzehntelang bewährte Intelligenzdiagnostik anschließt und vollständig auf Leistungsdaten statt auf Selbstauskünfte setzt.

Tabelle 3: Bekannte internationale Fremdeinschätzungs- und Leistungstestverfahren und ihre dimensionale Struktur

Verfahren	Quelle	Format / Items	Dimensionale Struktur
AI Literacy Test	Hornberger et al. (2023)	30 MC-Items (+1 Anker)	Inhaltlich breit (Konzepte, ML, Anwendung, Gesellschaft, Ethik); empirisch weitgehend eindimensional
AILIT-S	Hornberger et al. (2025)	10 MC-Items (Kurzform)	Eindimensionaler Gesamtscore; nur für Gruppenvergleiche/ Monitoring
GLAT (Generative AI Literacy Assessment Test)	Jin et al. (2025)	20 szenario-basierte MC-Items	Kennen & Verstehen, Nutzen & Anwenden, Bewerten & Gestalten, Ethik (nach Ng et al., 2021)
AICOS (AI Competency Objective Scale)	Markus et al. (2025)	51 MC-Items (Kurzform 18)	KI verstehen, anwenden, erkennen, erzeugen/gestalten, KI-Ethik, Generative KI (modular)
SAIL4ALL	Soto-Sanfiel et al. (2025)	Wahr/Falsch-Wissenstest	4 Themen (Was ist KI? / Was kann KI? / Wie funktioniert KI? / Wie sollte KI genutzt werden?), empirisch 5 Scores
A-Faktor-Aufgabenbatterie	Li et al. (2025)	18 Performanzaufgaben	Hierarchisch: Generalfaktor (A-Factor) über 4 Dimensionen (Kommunikation mit KI, kreative Ideengenerierung, Inhaltsbewertung, Problemzerlegung/Kollaboration)

Ergänzend sei erwähnt, dass internationale Organisationen parallel zur Testentwicklung normative Rahmenwerke vorgelegt haben – etwa die KI-Kompetenzrahmen der UNESCO für Lernende und Lehrkräfte oder das gemeinsam von der OECD und der Europäischen Kommission initiierte AILit-Rahmenwerk für den Schulbereich. Solche Rahmenwerke definieren Soll-Kompetenzen, sind aber selbst keine Messverfahren; sie beschreiben den Inhaltsraum, den psychometrische Instrumente anschließend messbar machen müssen.

5. Die wichtigsten Verfahren im nationalen Umfeld (DACH-Raum)

Wer die internationale Literatur aufmerksam gelesen hat, dem wird ein bemerkenswerter Befund nicht entgangen sein: Ein erheblicher Teil der international führenden KI-Kompetenz-Verfahren stammt aus dem deutschsprachigen Raum (Deutschland, Österreich, Schweiz; im Folgenden kurz DACH-Raum). Die Grenze zwischen „international" und „national" verläuft hier also weniger zwischen den Instrumenten selbst als zwischen Publikationssprache und Einsatzkontext: Die führenden deutschsprachigen Forschungsgruppen publizieren ihre Verfahren auf Englisch in internationalen Fachzeitschriften, entwickeln und validieren sie aber ganz überwiegend an deutschen Stichproben – und stellen vielfach deutschsprachige Versionen bereit.

An erster Stelle ist erneut die **MAILS** zu nennen (Carolus et al., 2023). Sie wurde am Lehrstuhl für Psychologie der Mensch-Computer-Interaktion und Medienpsychologie der Universität Würzburg entwickelt und an einer deutschen Stichprobe validiert; die Items liegen auf Deutsch und Englisch vor. Im DACH-Raum ist die MAILS das De-facto-Standardinstrument, wenn KI-Kompetenz per Selbsteinschätzung erhoben werden soll – sie wird in Unternehmensbefragungen, Hochschulstudien und Evaluationen von Weiterbildungsprogrammen gleichermaßen eingesetzt. Inzwischen existiert auch eine validierte Kurzversion, was den Einsatz in betrieblichen Befragungen zusätzlich erleichtert. Ihre bereits beschriebene vierteilige Struktur – KI-Literacy im engeren Sinne, KI-Selbstwirksamkeit, KI-Selbstkompetenz und KI-Gestaltung – ist zugleich die differenzierteste Selbstberichtsstruktur im deutschsprachigen Raum.

Das zweite deutschsprachige Referenzverfahren für die Selbsteinschätzung ist die **SNAIL** (Laupichler et al., 2023) aus der medizindidaktischen Forschung der Universität Bonn. Mit ihren drei Dimensionen – technisches Verständnis, kritische Bewertung, praktische Anwendung – ist sie schlanker aufgebaut als die MAILS und wurde ursprünglich für die medizinische Aus- und Weiterbildung entwickelt, inzwischen aber weit darüber hinaus adaptiert und international in mehreren Sprachfassungen weitervalidiert.

Auf der Seite der objektiven Verfahren dominiert im DACH-Raum die Münchner Arbeitsgruppe um Claudia Nerdel an der TUM School of Social Sciences and Technology: Der **AI Literacy Test** (Hornberger et al., 2023) und seine Kurzform **AILIT-S** (Hornberger et al., 2025) wurden an deutschen Studierendenstichproben entwickelt bzw. unter Beteiligung deutscher Stichproben international validiert und sind die derzeit am besten dokumentierten deutschsprachig verfügbaren Leistungstests für allgemeine KI-Literacy. Ebenfalls aus dem DACH-Raum – und zwar erneut aus Würzburg – stammt mit der **AICOS** (Markus et al., 2025) das aktuell umfassendste objektive Verfahren mit seiner sechsdimensionalen Struktur; es steht in deutscher und englischer Sprache zur Verfügung und wurde explizit mit dem Anspruch entwickelt, die Schwächen reiner Selbstberichtsskalen zu überwinden.

Eine Sonderstellung nimmt das bereits vorgestellte **AIEDTEC-Modell** der Frankfurter Gruppe um Andreas Frey ein (Frey et al., 2026). Es ist das derzeit elaborierteste deutschsprachige *Theoriemodell* zur KI-Kompetenz im Bildungskontext: Es verbindet systemisches Denken (Systems Thinking) als übergreifende Denkhaltung, vier Wissensbereiche in Anlehnung an das TPACK-Modell, sechs kognitive Prozessniveaus nach der revidierten Bloom-Taxonomie und einen vierphasigen Kompetenzentwicklungsverlauf zu einer Matrix, aus der sich systematisch Testaufgaben ableiten lassen. Wichtig für die korrekte Einordnung ist allerdings: Der im selben Beitrag validierte achtitemige Evaluationsfragebogen (fünfstufige Likert-Skala; drei Dimensionen in Anlehnung an das Kirkpatrick-Modell der Trainingsevaluation: Reaktion, Lernen, Verhalten; sehr gute Modellpassung mit CFI = .990 und RMSEA = .046 sowie Cronbachs Alpha = .87 – Kennwerte, die die Zuverlässigkeit und Strukturgüte des Fragebogens belegen) misst nicht die individuelle KI-Kompetenz einer Person, sondern die Qualität von Lehrveranstaltungen, die diese Kompetenz vermitteln sollen. Die Entwicklung eines echten Kompetenztests auf Basis des AIEDTEC-Modells benennen die Autoren selbst als nächsten Forschungsschritt. Für Unternehmen ist das Modell dennoch schon heute wertvoll – als Blaupause dafür, wie ein systematischer Kompetenzaufbau vom Grundlagenwissen bis zur eigenständigen Gestaltung KI-gestützter Prozesse gestuft werden kann.

Mit Blick auf die Arbeitswelt verdient schließlich das **AIComp-Rahmenmodell** („Artificial Intelligence Competences") besondere Beachtung (Ehlers et al., 2024). Es wurde vom NextEducation-Forschungsteam um Ulf-Daniel Ehlers an der Dualen Hochschule Baden-Württemberg (DHBW) im Kontext des KI-Campus entwickelt und versteht sich als erste vollständig empirisch-quantitativ konstruierte Future-Skills-Modellierung der KI-Kompetenz – also als Modell jener Zukunftskompetenzen, die Menschen für eine durch KI geprägte Lebens- und Arbeitswelt benötigen. Methodisch wurde in mehreren Schritten vorgegangen: Auf eine Literaturanalyse und qualitative Interviews mit Studierenden, Lehrenden, Wirtschaftsakteuren und Zivilgesellschaft folgte eine quantitative Onlinebefragung von Erwerbstätigen mit 1.644 auswertbaren Fragebögen, deren 36 Kompetenzitems per Hauptkomponentenanalyse (einem faktorenanalytischen Verfahren zur Verdichtung vieler Einzelitems auf wenige zugrundeliegende Komponenten) ausgewertet wurden; die Zwölf-Komponenten-Lösung klärte dabei 70 Prozent der Varianz auf. Das resultierende Modell ordnet zwölf Kompetenzfelder drei übergeordneten Kompetenzbereichen zu: dem Bereich Arbeit und Aufgabe (Aktivitäts- und Umsetzungskompetenz, Systemdesignkompetenz, kreative Problemlösekompetenz, kritische digitale Kompetenz, Entscheidungskompetenz), dem Bereich persönliche Entwicklung (Selbstwirksamkeit, kritisches Denken, aktive Steuerungsfähigkeit, Selbstbestimmtheit) und dem Bereich soziales Umfeld und Organisation (ethische Kompetenz, Kooperationskompetenz, Kommunikationskompetenz). Damit spannt AIComp den Bogen deutlich weiter als die klassischen AI-Literacy-Skalen: Neben funktionalen Facetten stehen ausdrücklich personale und sozial-organisationale Kompetenzfelder – eine Perspektive, die konzeptionell an die Meta-Kompetenzen der MAILS

anschließt und für betriebliche Personal- und Organisationsentwicklung unmittelbar anschlussfähig ist. Zugleich gilt die in Kapitel 3 dargelegte Einschränkung: AIComp beruht auf Selbsteinschätzungen mit Zustimmungsskalen und misst damit wahrgenommene Kompetenz und Bedeutungszuschreibungen, keine objektive Leistung. Sein besonderer Wert liegt deshalb weniger in der Individualdiagnostik als in der inhaltlichen Landkarte dessen, was eine KI-geprägte Arbeitswelt jenseits der reinen Werkzeugbeherrschung verlangt.

Jenseits der akademischen Verfahren hat sich im DACH-Raum eine breite Praxislandschaft entwickelt: Weiterbildungsanbieter, Prüforganisationen und Plattformen wie der vom Bundesministerium für Bildung und Forschung geförderte KI-Campus bieten Selbsttests, Wissenschecks und Zertifikatskurse an, und im Nachgang zu Artikel 4 der KI-Verordnung sind zahlreiche kommerzielle „KI-Führerschein"- und Compliance-Schulungsformate entstanden. Aus psychometrischer Sicht ist hier allerdings Zurückhaltung geboten: Die wenigsten dieser Praxisinstrumente sind nach testtheoretischen Standards konstruiert oder unabhängig validiert. Wer den Kompetenzstand seiner Organisation belastbar erfassen will, sollte auf die wissenschaftlich geprüften Verfahren zurückgreifen oder eigene Verfahren nach deren Vorbild professionell entwickeln lassen.

Tabelle 4: Die wichtigsten Verfahren mit Ursprung im DACH-Raum im Überblick

Verfahren	Institution	Messzugang	Dimensionale Struktur
MAILS	Universität Würzburg	Selbsteinschätzung (Likert)	4 Bereiche: AI Literacy (4 Facetten), KI-Selbstwirksamkeit (2), KI-Selbstkompetenz (2), KI-Gestaltung
SNAIL	Universität Bonn	Selbsteinschätzung (Likert)	3 Dimensionen: Technisches Verständnis, Kritische Bewertung, Praktische Anwendung
AI Literacy Test / AILIT-S	TU München	Objektiver Wissenstest (MC)	Inhaltlich breit, empirisch weitgehend eindimensional; Kurzform nur für Gruppenvergleiche
AICOS	Universität Würzburg	Objektiver Wissenstest (MC)	6 Subkompetenzen: Verstehen, Anwenden, Erkennen, Erzeugen/Gestalten, Ethik, Generative KI
AIEDTEC-Modell (+ Evaluationsfragebogen)	Goethe-Universität Frankfurt / DIPF (Leibniz-Institut für Bildungsforschung und Bildungsinformation)	Theorierahmen; Fragebogen zur Lehrevaluation (Likert)	Modell: Systems Thinking × 4 Wissensbereiche × 6 kognitive Prozesse × 4 Entwicklungsphasen; Fragebogen: Reaktion, Lernen, Verhalten
AIComp	DHBW / NextEducation (KI-Campus)	Selbsteinschätzung (Zustimmungsskalen, 36 Items)	3 Kompetenzbereiche (Arbeit & Aufgabe; persönliche Entwicklung; soziales Umfeld & Organisation) mit 12 empirisch ermittelten Kompetenzfeldern

Für Österreich und die Schweiz gilt: Eigenständige, psychometrisch publizierte KI-Kompetenztests von vergleichbarer Reichweite sind bislang rar; Forschung und Praxis greifen dort überwiegend auf die genannten deutschen bzw. internationalen Verfahren zurück, teils in adaptierter Form. Pädagogische Hochschulen in der Schweiz und österreichische Bildungseinrichtungen arbeiten derzeit vor allem an Rahmenmodellen und Curricula; die Instrumentenentwicklung im engeren Sinne konzentriert sich im deutschsprachigen Raum sichtbar auf die genannten Standorte Würzburg, München, Bonn und Frankfurt.

6. Bilanz: Welche dimensionale Struktur hat gegenwärtig die größte Verbreitung?

Fasst man die internationale und die deutschsprachige Verfahrenslandschaft zusammen, lässt sich die Ausgangsfrage dieses Artikels – welche dimensionale Struktur der KI-Kompetenz sich derzeit durchsetzt – erstaunlich klar beantworten, sofern man zwei Ebenen sauber trennt: die *inhaltliche* Ebene (welche Kompetenzbereiche werden unterschieden?) und die *empirisch-statistische* Ebene (welche Faktorenstruktur zeigen die Daten?).

Auf der inhaltlichen Ebene hat sich ein **Vier-Dimensionen-Kern** als der mit Abstand verbreitetste gemeinsame Nenner etabliert. Er geht maßgeblich auf das Rahmenmodell von Ng et al. (2021) zurück und findet sich – in leicht variierender Benennung – in der Mehrzahl aller aktuellen Verfahren: erstens **KI kennen und verstehen** (deklaratives Grundlagen- und Funktionswissen), zweitens **KI nutzen und anwenden** (prozedurale Anwendungskompetenz), drittens **KI-Ergebnisse bewerten** (kritische Evaluation von Qualität, Verlässlichkeit und Grenzen) und viertens **KI verantwortungsvoll und ethisch nutzen** (ethische, rechtliche und gesellschaftliche Verantwortung). Der GLAT operationalisiert exakt diese vier Dimensionen; die AICOS, die MAILS und zahlreiche berufsspezifische Skalen enthalten sie als Teilmenge. Rahmenwerke wie die der UNESCO bauen erkennbar auf demselben Grundgerüst auf. Neuere Verfahren ergänzen diesen Kern typischerweise um bis zu drei weitere Bereiche: das **Erkennen** von KI und KI-generierten Inhalten, das **Gestalten bzw. Erzeugen** von KI-Lösungen sowie – als jüngste Entwicklung – eine eigene **Generative-KI- bzw. Prompting-Dimension** und Facetten der **kontinuierlichen Kompetenzaktualisierung**. Dass sich der Blick dabei zunehmend über den funktionalen Kern hinaus weitet, zeigt im deutschsprachigen Raum exemplarisch das AIComp-Modell (Ehlers et al., 2024): Dort treten – ähnlich wie in den Meta-Kompetenzen der MAILS – Selbstwirksamkeit, Selbststeuerung, Kooperation und Kommunikation als eigenständige Kompetenzfelder einer KI-geprägten Arbeitswelt neben die klassischen Wissens- und Anwendungsdimensionen.

Auf der empirisch-statistischen Ebene ist das Bild differenzierter – und für die Praxis hochrelevant. Über nahezu alle objektiven Tests hinweg zeigt sich, dass die inhaltlich unterschiedenen Dimensionen hoch miteinander korrelieren und häufig ein starker Generalfaktor dominiert: Wer viel über KI weiß, kann sie meist auch besser anwenden und kritischer bewerten. Der Münchner AI Literacy Test erwies sich als weitgehend eindimensional; beim GLAT dominiert trotz vierdimensionaler Inhaltsstruktur ein Gesamtfaktor; die Tsinghua-Gruppe hat den Generalfaktor mit ihrem A-Factor-Ansatz sogar zum Programm erhoben (Li et al., 2025); und neuere berufsspezifische Skalen – etwa eine 2026 für Pflegefachkräfte publizierte Generative-AI-Literacy-Skala – modellieren ihre Facetten explizit unter einem übergeordneten Faktor zweiter Ordnung. Inhaltsdimensionen und latente Faktoren fallen also nicht zusammen: Ein Test kann vier oder sechs Inhaltsbereiche systematisch abdecken und trotzdem im Kern eine einzige zugrundeliegende Fähigkeit messen.

Die Konsequenz, auf die sich die Forschung 2025/2026 erkennbar zubewegt, ist ein **hierarchisches Strukturmodell**: ein allgemeiner KI-Kompetenzfaktor an der Spitze, darunter vier bis sechs korrelierte Kernfacetten, ergänzt um rollen-, branchen- oder technologiespezifische Module, die als Inhaltsbausteine – nicht zwingend als eigenständige latente Dimensionen – geführt werden. Dieses Modell verbindet das Beste beider Welten: Der Gesamtwert erlaubt robuste Vergleichs- und Verlaufsaussagen, die Facettenprofile liefern diagnostisch nutzbare Ansatzpunkte für gezielte Entwicklung, und die Modulbauweise hält das System offen für den technologischen Wandel.

Tabelle 5: Der dimensionale Konsenskern und seine Verbreitung in aktuellen Verfahren

Dimension	Inhalt	Beispiele für Verfahren, die sie explizit ausweisen
KI kennen & verstehen	Grundlagen-, Funktions- und Konzeptwissen	GLAT, AICOS, SAIL4ALL, MAILS, SNAIL
KI nutzen & anwenden	Zielgerichtete praktische Anwendung und Interaktion	GLAT, AICOS, MAILS, SNAIL
KI-Ergebnisse bewerten	Kritische Prüfung von Qualität, Verlässlichkeit, Grenzen	GLAT (Evaluate & Create), AICOS (implizit), MAILS, SNAIL, A-Faktor-Batterie (Inhaltsbewertung)
Verantwortungsvolle & ethische Nutzung	Ethik, Recht, gesellschaftliche Folgen	GLAT, AICOS, MAILS
Erweiterung: KI erkennen	Identifikation von KI-Systemen und KI-generierten Inhalten	AICOS, MAILS
Erweiterung: KI gestalten / erzeugen	Entwicklung, Konfiguration, Workflow-Integration	AICOS, MAILS (KI-Gestaltung), A-Faktor-Batterie (Kollaboration/Problemzerlegung)
Erweiterung: Generative KI / Prompting	Spezifischer Umgang mit generativen Systemen	AICOS (eigenes Modul), GLAT (durchgängiger Gegenstand)

Als Zwischenfazit lässt sich festhalten: Die größte Verbreitung hat gegenwärtig der vierdimensionale Inhaltskern „Verstehen – Anwenden – Bewerten – Verantwortungsvoll nutzen", zunehmend erweitert um Erkennen, Gestalten und generative KI; psychometrisch setzt sich zugleich die Erkenntnis durch, dass diese Inhaltsdimensionen am überzeugendsten hierarchisch – als Facetten unter einem Generalfaktor – modelliert werden.

7. Fazit und Ausblick: Wie Unternehmen und Organisationen in den kommenden zwei bis drei Jahren vorgehen sollten

Der Blick auf den Forschungsstand 2025/2026 ergibt ein doppeltes Bild. Einerseits ist das Feld deutlich reifer geworden: Es existiert ein belastbarer inhaltlicher Konsenskern von vier Dimensionen, es liegen erstmals mehrere gut konstruierte objektive Tests vor, und zentrale methodische Einsichten – die begrenzte Aussagekraft von Selbstauskünften, die Dominanz eines Generalfaktors, der Unterschied zwischen Inhaltsblueprint und Faktorenstruktur – sind empirisch solide abgesichert. Andererseits bleibt Wesentliches offen: Ob generative KI dauerhaft eine eigene Dimension bildet oder zur Querschnittstechnologie aller Facetten wird, ist ungeklärt; Messinvarianz über Berufsgruppen, Sprachen und Kulturen ist erst für wenige Verfahren nachgewiesen; verbindliche Kompetenzniveaus und Schwellenwerte (Cut-Scores) für Zertifizierungen fehlen; und wie gut Testwerte reale Arbeitsleistung außerhalb von Hochschulkontexten vorhersagen, ist erst ansatzweise untersucht. Eine „finale" Definition der KI-Kompetenz wird es auf absehbare Zeit nicht geben – das Konstrukt wird sich mit der Technologie weiterentwickeln, so wie sich auch das Verständnis von Lese- oder Medienkompetenz historisch gewandelt hat.

Für Unternehmen und Organisationen ist diese Offenheit jedoch kein Grund zum Abwarten – im Gegenteil: Wer auf die endgültige Definition wartet, verstößt schon heute gegen den Geist von Artikel 4 der KI-Verordnung und verliert wertvolle Zeit im Kompetenzaufbau. Aus dem dargestellten Forschungsstand lässt sich ein pragmatisches, wissenschaftlich fundiertes Vorgehen für die kommenden zwei bis drei Jahre ableiten, das sich in drei Phasen gliedern lässt.

In der **ersten Phase (die nächsten sechs Monate)** sollten Organisationen eine strukturierte Bestandsaufnahme vornehmen. Grundlage sollte der vierdimensionale Konsenskern sein – Verstehen, Anwenden, Bewerten, verantwortungsvolle Nutzung –, ergänzt um die für das eigene Geschäft relevanten Module (generative KI, branchenspezifische Anwendungen, unternehmensinterne Richtlinien). Für ein

erstes Lagebild ist die Kombination aus einer validierten Selbsteinschätzungsskala (etwa MAILS oder SNAIL) und einem kurzen objektiven Wissenscheck (etwa nach dem Vorbild von AILIT-S oder ausgewählten AICOS-Modulen) sinnvoll: Die Selbsteinschätzung liefert Akzeptanz-, Motivations- und Selbstbilddaten, der Wissenstest den Realitätsabgleich. Schon die Differenz zwischen beiden Werten ist diagnostisch wertvoll, denn sie zeigt, wo Selbstüberschätzung die größten Risiken birgt – typischerweise dort, wo intensiv mit KI gearbeitet, aber wenig geprüft wird.

In der **zweiten Phase (sechs bis achtzehn Monate)** geht es um den systematischen Kompetenzaufbau und dessen Verankerung in der Personalentwicklung. Hier bewährt sich die Entwicklungslogik des AIEDTEC-Modells als Blaupause: vom Grundlagenwissen (Erinnern, Verstehen) über die Anwendung und Analyse in bekannten Situationen hin zur Bewertung von Qualität und Risiken und schließlich zur eigenständigen Gestaltung KI-gestützter Arbeitsprozesse. Rollen sollten differenziert werden – nicht jede Mitarbeiterin braucht Gestaltungskompetenz, aber jede braucht Bewertungs- und Verantwortungskompetenz. Trainings sollten konsequent fertigkeitsorientiert angelegt sein (Übung an realen Arbeitsaufgaben, Feedback, Simulationen statt reiner Wissensvermittlung), und die Wirksamkeit sollte zweistufig evaluiert werden: kurzfristig über Reaktions- und Lernmaße, mittelfristig über objektive Kompetenz- und Verhaltensindikatoren.

In der **dritten Phase (achtzehn bis sechsunddreißig Monate)** sollten Organisationen ihr Kompetenzmanagement professionalisieren und auf Dauer stellen. Dazu gehört erstens ein wiederkehrendes Monitoring mit stabilen Ankeritems, das Kompetenzverläufe über die Zeit vergleichbar macht, während technologie- und toolspezifische Items regelmäßig erneuert werden; zweitens der Aufbau bzw. Einkauf objektiver, idealerweise szenario- und simulationsbasierter Verfahren für kritische Rollen und Entscheidungen (Personalauswahl, Freigabe sensibler KI-Anwendungen, interne Zertifizierungen); drittens die Beobachtung der Standardisierungsentwicklung, denn es ist absehbar, dass sich aus Wissenschaft, Normung und Regulierungspraxis in den kommenden Jahren anerkannte Referenzrahmen und möglicherweise zertifizierbare Prüfstandards für KI-Kompetenz herausbilden werden. Organisationen, die bis dahin eine eigene, datenbasierte Kompetenzarchitektur aufgebaut haben, werden solche Standards ohne Bruch adaptieren können.

Quer zu allen drei Phasen liegt eine Grundsatzentscheidung, die sich aus der Forschung eindeutig ableiten lässt: **Selbsteinschätzung ergänzt, aber ersetzt niemals objektive Messung.** Wo es um belastbare Aussagen geht – gegenüber Aufsichtsbehörden, in Personalentscheidungen, bei der Freigabe risikobehafteter KI-Anwendungen –, müssen demonstrierte Kompetenz und dokumentierte Prüfung den Kern bilden. Und ebenso klar gilt: KI-Kompetenz ist kein einmalig zu erwerbendes Zertifikat, sondern eine dynamische Kompetenz, deren Halbwertszeit kürzer ist als bei fast jeder anderen beruflichen Qualifikation. Die Fähigkeit und Bereitschaft zur kontinuierlichen Aktualisierung – von neueren Verfahren bereits als eigene Facette modelliert – dürfte sich in den kommenden Jahren als die vielleicht wichtigste Teilkompetenz überhaupt erweisen.

Die dimensionale Struktur der KI-Kompetenz ist damit mehr als eine akademische Fingerübung. Sie ist die Landkarte, mit der Organisationen entscheiden können, was sie messen, was sie entwickeln und was sie nachweisen wollen. Diese Landkarte ist heute – anders als noch vor drei Jahren – gut genug, um verantwortungsvoll damit zu navigieren. Wer sie nutzt, verschafft sich einen Vorsprung, der mit jedem Monat wächst.

Literaturverzeichnis

Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS – Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models

and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans, 1*(2), 100014. https://doi.org/10.1016/j.chbah.2023.100014

Ehlers, U.-D., Lindner, M., & Rauch, E. (2024). *AIComp – Future Skills für eine von KI beeinflusste Lebens- und Arbeitswelt. Forschungsbericht 2: Empirische Konstruktion & Beschreibung des Kompetenzmodells AIComp*. NextEducation, Duale Hochschule Baden-Württemberg. https://aicomp.org/downloads/AIComp_Part_2_Kompetenzmodell_final.pdf

Frey, A., Schenk, C., Weiß, L., König, C., Ramesh, V., Fabriz, S., Drachsler, H., & Horz, H. (2026). A theoretical framework and evaluation instrument for the competence to use artificial intelligence and digital technology in educational processes. *Systems, 14*(5), 583. https://doi.org/10.3390/systems14050583

Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence, 5*, 100165. https://doi.org/10.1016/j.caeai.2023.100165

Hornberger, M., Bewersdorff, A., Schiff, D. S., & Nerdel, C. (2025). Development and validation of a short AI literacy test (AILIT-S) for university students. *Computers in Human Behavior: Artificial Humans*, 100176. https://doi.org/10.1016/j.chbah.2025.100176

Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2025). GLAT: The Generative AI Literacy Assessment Test. *Computers and Education: Artificial Intelligence, 9*, 100436. https://doi.org/10.1016/j.caeai.2025.100436

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "Scale for the assessment of non-experts' AI literacy" – An exploratory factor analysis. *Computers in Human Behavior Reports, 12*, 100338. https://doi.org/10.1016/j.chbr.2023.100338

Li, N., Deng, W., & Chen, J. (2025). *From G-factor to A-factor: Establishing a psychometric framework for AI literacy* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2503.16517

Lintner, T. (2024). A systematic review of AI literacy scales. *npj Science of Learning, 9*, 50. https://doi.org/10.1038/s41539-024-00264-4

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (S. 1–16). ACM. https://doi.org/10.1145/3313831.3376727

Markus, A., Carolus, A., & Wienrich, C. (2025). Objective measurement of AI literacy: Development and validation of the AI Competency Objective Scale (AICOS). *Computers and Education: Artificial Intelligence, 9*, 100485. https://doi.org/10.1016/j.caeai.2025.100485

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Soto-Sanfiel, M. T., Angulo-Brunet, A., & Lutz, C. (2025). The scale of artificial intelligence literacy for all (SAIL4ALL): Assessing knowledge of artificial intelligence in all adult populations. *Humanities and Social Sciences Communications, 12*, 1618. https://doi.org/10.1057/s41599-025-05978-3

Rechtsquelle: Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (KI-Verordnung / AI Act), Artikel 4 (KI-Kompetenz). Amtsblatt der Europäischen Union, L-Reihe, 12.07.2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

© Lieberei Consulting – Blog-Artikel Nr. 47. Alle Angaben nach bestem Wissen auf Basis des Forschungsstands Juli 2026; keine Rechtsberatung.