

SEVAL

Schweizerische Evaluationsgesellschaft
Société suisse d'évaluation
Società svizzera di valutazione

SEVAL – AG KI

Wissensstände, Überlegungen und Tätigkeiten der AG KI

Kerstin Schwarz
FMSystO GmbH
Turbenweg 1
CH-3073 Gümligen
+41 76 831 5554
info@fmsysto.com



Carole Probst
econcept AG
Gerechtigkeitsgasse 20
CH-8001 Zürich
+41 44 286 7575
info@econcept.ch



Präsentation an die Teilnehmenden zur Sitzung vom 17.06.2025

Steffisburg, 16. Juni 2025

Inhaltsverzeichnis

0	Vorstellung der Arbeitsgruppe KI (SEVAL AG KI)	3
1	Anwendungserfahrung (KI), Überlegungen und Tätigkeiten	4
2	Ausblick	12

Entstehungsgeschichte und Struktur

Überblick: Arbeitsgruppe KI



SEVAL

Schweizerische Evaluationsgesellschaft
Société suisse d'évaluation
Società svizzera di valutazione

Allgemeine Informationen

- 2024: Die Arbeitsgruppe «Künstliche Intelligenz» (AG KI) wurde gegründet, um eine gemeinsame Herangehensweise und koordinierte Initiative zum Thema KI in der Evaluation zu entwickeln. Dazu wurde durch den Vorstand eine Interessensbekundung ausgerufen.
- 2025.02.19: De 1. gemeinsamen Sitzung der AG KI: Erhebung des aktuellen Stands zu Erfahrung und Wissen im Bereich «Künstliche Intelligenz» in der Arbeitsgruppe u. Organisation

Zahlen & Daten

- 22 Mitglieder (gesamthft)
- 2 Co-Sprecherinnen:



Carole Probst

econcept AG
Partnerin / Mitglied der
Geschäftsleitung

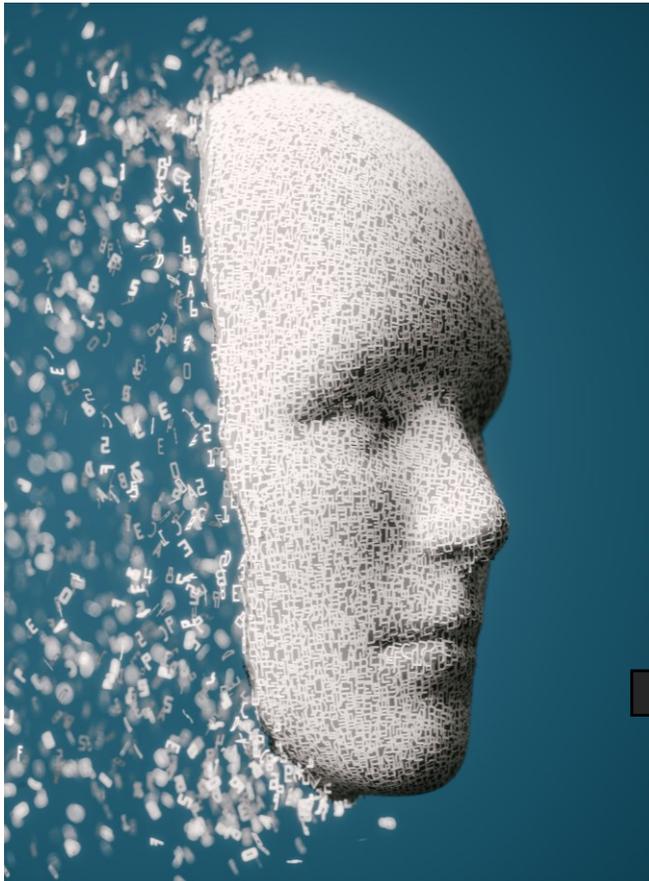


Kerstin Schwarz

FMSystO GmbH
Geschäftsleitung

Bestandesaufnahme von Wissen und Erfahrungen in der AG KI

Umfrage und Grundlage für zukünftige Aktivitäten in der AG KI



Themen der Umfrage

1. Theoretische Erfahrungen mit KI
2. Praktische Erfahrungen mit KI
3. Anwendungs- und Praxisbeispiel
4. Veränderung der Arbeitsweise
5. Interessen
6. Organisation und Termine der AG KI
7. Sonstiges

Details zur Durchführung



10.03.2025 – 28.03.2025 | 18 Tage



22 Personen im Verteiler



15 Teilnahmen

Relevante KI-Konzepte in der Evaluationspraxis

Anwendungen wie Textauswertung, Explainable AI und generative Modelle

Q 1.2: Beschreibe bitte kurz ein KI-Konzept, das für Dich besonders relevant für die Evaluationspraxis ist.

Beschreibung	Häufigkeit
Auswertung von Text-Daten mittels LLM: wie kann man Pipelines aufbauen, denen man "vertrauen" kann, um Daten auszuwerten; Beschleunigung qualitativer Auswertung, "Zweitmeinung", mehr Auseinandersetzung mit den Quellen in kürzerer Zeit	1
Ein Konzept von KI, das für die Evaluierungspraxis besonders relevant ist, ist sein Einsatz im gesamten Prozess, einschließlich der Entwicklung von Evaluierungsfragen, der Modellierung von Auswirkungen und der Analyse von Daten. Auch die damit verbundenen Risiken, insbesondere die Frage des Schutzes personenbezogener Daten, interessieren mich sehr.	1
Explainable AI (XAI): Entscheidungsprozesse und Nachvollziehbarkeit können transparent gemacht werden, hilft BIAs zu identifizieren, erreicht die geforderte Genauigkeit	1
KI kann besonders nützlich sein, um große Datenmengen schnell zu verarbeiten. Beispielsweise um Dokumente zusammenzufassen, zu identifizieren, welches Dokument welche Informationen enthält oder um zu ermitteln, wie viele Personen in einer Vielzahl von Interviewprotokollen eine Meinung geäußert haben.	1
Ich benutze MAXQDA Software für Datenanalyse und jetzt gibt es Talwind. Man kann grosse Daten zu kurzen zusammenfassen, ohne alles zuerst zu lesen und einen Rapport schreiben. Was verlieren wir damit?	1
<ul style="list-style-type: none"> - KI-System zur Datenaggregation und -verwaltung (Text oder verschlüsselt); - Übersetzung oder in bestimmten Fällen sogar Verbesserung von Formulierungen; - Visualisierung 	1
Für meine Evaluationspraxis ist generative learning besonders relevant, weil es einfach zu benutzen ist und vielseitig ist. Ich benutze Perplexity.	1
Generative KI (aber nur, weil ich es bisher bereits in der Praxis angewendet habe und die anderen KI-Konzepte nicht wirklich kenne)	1
Entwicklungen von Sprachmodellen und autonom lernende und handelnde KI. Nützlich in allen Phasen der Evaluation (Literaturübersichten, Qualitätssicherung, statistische Analysen).	1

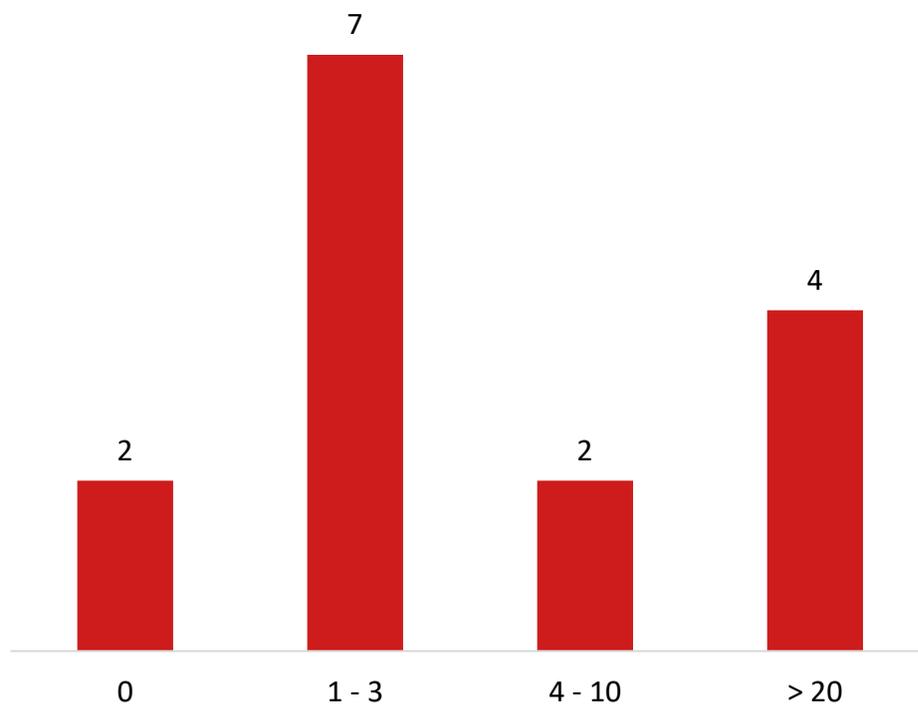
Nennungen: 9

Fazit: Allg. wird Generative KI als vielseitig und unterstützend bei Textauswertung, Übersetzung und Visualisierung wahrgenommen. Es gibt einzelne Betonungen: LLMs zur Beschleunigung der qualitativen Textauswertung oder Explainable AI für die Transparenz von Entscheidungsprozessen. Eine besondere Relevanz eines spezifischen Konzeptes scheint nicht zu existieren.

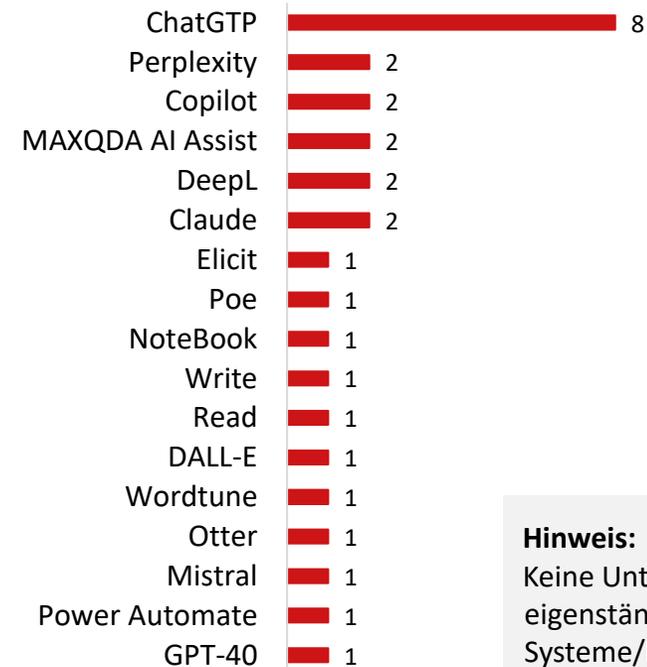
Nutzung spezifischer KI-Tools in der Praxis

Überblick über verwendete Plattformen, Technologien und Projekte

Q. 2.1 Wie viele Evaluationsprojekte hast Du bereits durchgeführt, bei denen KI zum Einsatz kam?



Q. 2.1 Welche spezifischen KI-Tools, - Plattformen oder Technologien hast Du in Deiner praktischen Arbeit bereits verwendet?



Hinweis:
Keine Unterscheidung zwischen eigenständiger KI und KI-gestützte Systeme/KI-fähige Technologien.

Fazit: KI-Technologie wird in Evaluationsprojekten eingesetzt. Bei den spezifischen KI-Tools dominiert ChatGPT. Es ist anzunehmen, dass weniger gemeinsame Erfahrung bei der Nutzung anderer KI-Tools besteht.

Einsatzmöglichkeiten + Mehrwert von KI in Eval.-projekten

Bereiche, in denen KI eingesetzt wird, und die daraus resultierenden Vorteile für die Evaluation

Q. 2.3 In welchen Bereichen der Evaluation hast Du diese KI-Tools, -Plattformen oder -Technologien eingesetzt?



Q. 2.5 Welche Vorteile hast Du durch den Einsatz von KI in Deinen Evaluationsprojekten erfahren?

- **Zeit Gewinnung**
 - da ich mit Perplexity nicht wie bei Google den Inhalt der einzelnen Links analysieren muss.
 - Bei der Datenverwaltung/-aggregation
 - Visualisierung bestimmter Produktionen/Layouts/PDF
 - Einarbeitung in Themen, u. a. Gesundheitssystem Dänemark
 - Finden von Quellen
 - Unterstützung bei Sprachen
 - Literaturrecherche
 - Allg. Beschleunigung von Prozessen
 - Zeitersparnis beim Erstellen von Leitfäden (eher geringe Zeitersparnis)
 - Auswerten von Interviews (eher geringe Zeitersparnis)
- **Anpassung von Formulierungen**
 - Mit DeepL hin- und zurück übersetzen
 - Flüssigere Texte
- **Ideengenerierung**
 - Ideen generationen finde ich besonders gut und Clinical Decision support systems sind super und online forward triage tools in der Zeit der COVID haben sehr viel gebracht
- **Prüfung und Korrekturen**
 - Man benötigt keinen menschlichen "Sparring-Partner" bei Texten

Fazit: KI wird bereits in allen Phasen des Evaluationsprozesses (Planung, Durchführung, Berichterstattung) eingesetzt. KI scheint sich zunehmend als Werkzeug in der Evaluationspraxis zu etablieren. Sie wird überwiegend genutzt, um standardisierte und zeitaufwändige Aufgaben effizienter zu bearbeiten.

Herausforderung bei der Nutzung von KI

Datenschutz, Quellvalidität und Lösungsansätze

Q. 2.4 Beschreibe konkrete Herausforderungen, der Du bei der Anwendung von KI in einem Evaluationsprojekt begegnet bist und wie Du es gelöst hast

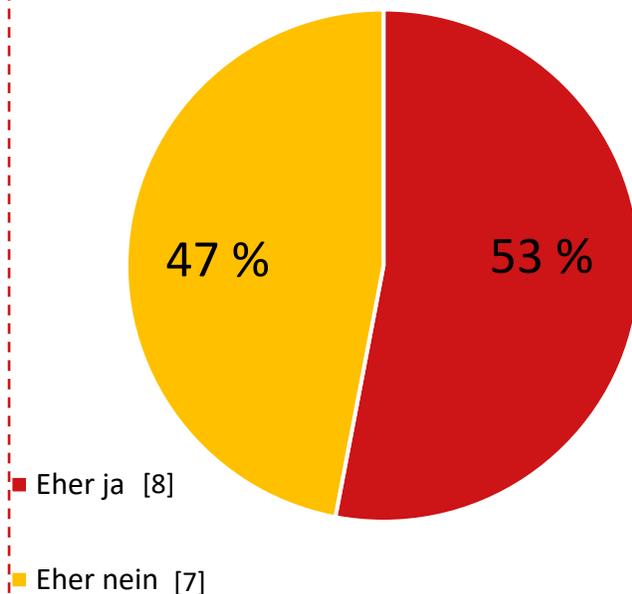
Beschreibung	Lösungsmöglichkeiten
ChatGTP erfindet Quellen in Literaturübersichten	<ul style="list-style-type: none"> - Leitfaden für qualitative Interviews erstellen: Beginn mit von KI erstelltem Entwurf anstatt einem weissen Blatt Papier - Transkription von Interviews: Ausprobieren verschiedener Tools für Interviews in Deutsch, Schweizerdeutsch und französisch. Durch eine automatische Transkription lassen sich später im Auswertungsprozess auch die gesuchten Textstellen einfach finden, falls das schriftliche Protokoll weniger ausführlich ist
Datenschutz	- noch nicht gelöst
Datenschutz , um vertrauliche Dokumente zu analysieren und generische Eingabeaufforderung zur Wiederverwendung in verschiedenen Projekten entwickeln zu können	- Zusammenarbeit mit Informatik-Experten zur Anpassung eines LLM
Datenschutz : Wo werden die Datengespeichert. Entscheidung, welche Informationen in das Tool eingefügt werden	- Jeder Fall wird individuell gelöst
ChatGPT hat nur eine bestimmte Art zuschreiben. Es fehlt die Kreativität	- Noch nicht gelöst
Auftragsformulierung : Eine Schlagwortsuche ist nicht ausreichend. Es erfordert Kommunikation wie mit Menschen. Mehrmalige Nach-Beschreibung kann erforderlich werden.	- Man beschreibt so lange, bis das Ergebnis zufriedenstellend ist
Überprüfung: Die Prüfung des Textes erfordert den gleichen Zeitaufwand	- Man investiert gleich zu Beginn mehr Zeit in die Auftragsbeschreibung
Wie kann man sinnvoll mit Halluzinationen umgehen?	- Einheiten der Abfragen sehr klein und sehr spezifisch machen (z.B. eine sehr spezifische Frage über einen kleinen Text) statt schwammige Abfragen über grossen Text-Korpus.
Am häufigsten benutze ich KI bei der Suche (Perplexity) und bei der Übersetzung (DeepL). Es ist mir auch schon passiert, Perplexity für Zusammenfassungen zu benutzen, wobei ich vorher den Text anonymisiert habe.	
In dieser Phase gibt es keine besonderen Herausforderungen, sondern eher eine Überlegung , wie ich es mit völliger Transparenz und innerhalb eines ethischen Rahmens in meinen Arbeitsprozess integrieren kann.	
Unterstützen bei der Entwicklung von Bewertungsfragen und Wirkungsmodellen .	

Fazit: Es ist anzunehmen, dass neue Kompetenzen und Ressourcen bei der Nutzung von KI in allen Schritten einer Evaluation benötigt werden: Diese beziehen sich auf *technisches, methodisch und ethisches Wissen und Fähigkeiten von Evaluierenden, um KI korrekt und verantwortungsvoll einzusetzen.*

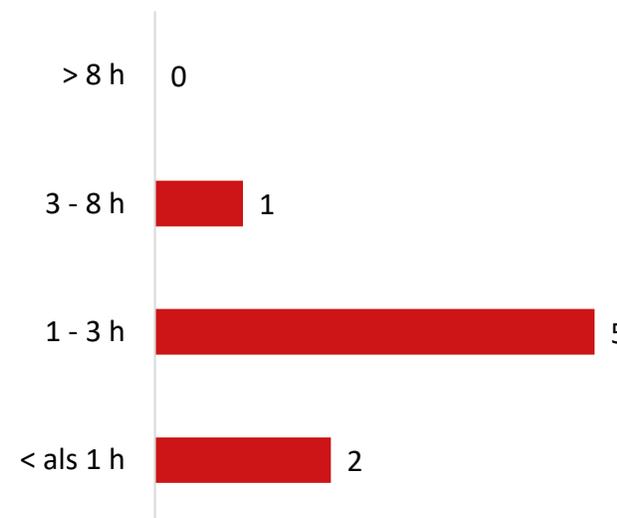
Neue Möglichkeiten durch KI in der Evaluation

Effizienzsteigerung, Automatisierung und innovative Ansätze

Q. 4.2 Hilft Dir KI, Zeit zu sparen?



Q. 4.2a Wenn ja, wie hoch schätzt Du die Zeitersparnis pro Woche?



Q. 4.2b Wofür nutzt Du die freigewordene Zeit in den Evaluationsprojekten?

- **Projektarbeit:** im entsprechenden Evaluationsprojekt, z.B. Schlussbericht, Grafiken etc.
- **Projektarbeit:** Ich setze sie entweder für den entsprechende Projekte oder für **andere Arbeits-projekte** ein, d.h. ich nehme nicht öfters frei!
- **Projektarbeit:** Durch die Zeitersparnis kann ich mehr **Dokumentenprüfungen** und Interviews durchführen oder sogar mehrere Mandate parallel wahrnehmen
- **Wissenstransfer:** Ich reflektiere und probiere mein eigenes Wissen und Kenntnisse zu integrieren
- Schwer zu sagen, ich habe noch nicht genug Erfahrung, um eine klare Antwort zu geben.
- **Freizeit:** und **Ehrenamt**
- **Projektarbeit:** Weitere Literaturrezensionen oder Arbeiten zu anderen Themen

Fazit: KI eröffnet Chancen u. verändert bestehende Arbeitspraktiken, bisher aber nur in Teilbereichen. Nutzer berichten über: effizientere Zeitverwendung, da KI-Standardprozesse (z. B. Transkription) übernimmt, mehr Zeit für komplexere Tätigkeiten, z. B. Projektmanagement, KI als Zweitmeinung zur Validierung qualitativer Trends, schnellere Dokumentenanalyse bei Literaturrecherche. Es entsteht ein hybrider Arbeitsmodus mit KI als Assistenzsystem.

Zukunft der Evaluation?

Automatisierung, neue Kompetenzen und strategische Innovationen durch den Einsatz von KI

Q. 4.2a Wie denkst Du, dass sich der Berufsalltag in der Evaluation durch den Einsatz der KI in der Zukunft verändern wird?

Effizienzsteigerung und Automatisierung von Routinetätigkeiten

- KI übernimmt Datenanalyse und Berichterstattung, wodurch mehr Zeit für strategische und kreative Aufgaben bleibt.

Veränderung der Rollen der Evaluator:innen

- Menschen bleiben für kritisches Denken
- Entscheidung und Interpretation unverzichtbar
- Der Fokus verschiebt sich auf die Prüfung und Validierung von KI-Ergebnissen

Neue technologischen und methodische Anforderungen

- Fachkräfte benötigen mehr technisches Wissen
- KI ermöglicht automatisierte Analysen und bessere Vergleichbarkeit von Projekten

Herausforderungen durch Standardisierung und Abhängigkeit

- Kreativität könnte durch standardisierte Verfahren eingeschränkt werden
- Es besteht das Risiko, dass kreative Prozesse durch KI-Lösungen verdrängt werden

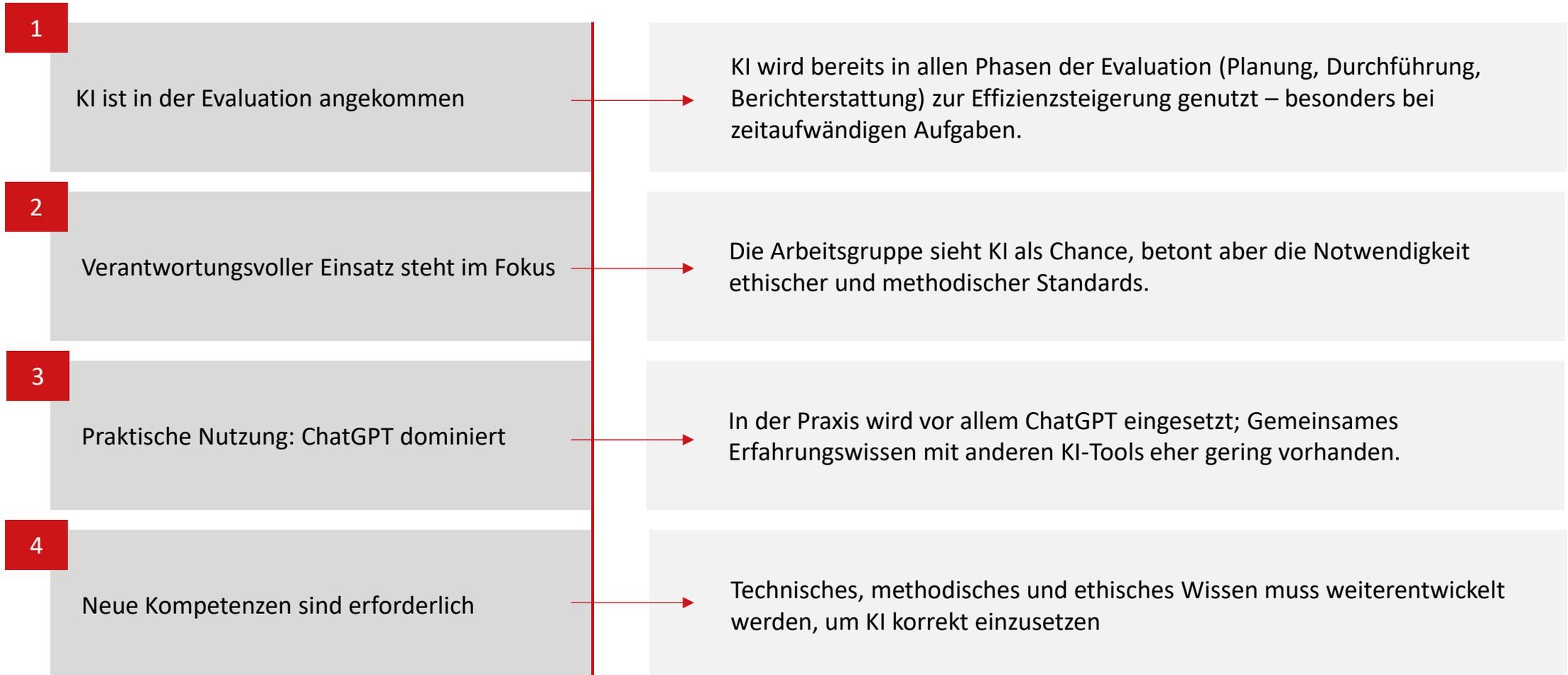
Langfristig: Transformation der Tätigkeit

- Der Fokus verschiebt sich auf komplexere Aufgaben und strategische Innovationen, während KI Routinearbeiten übernimmt

Fazit: KI wird vermutlich Routinetätigkeiten übernehmen, sodass mehr Zeit für kreative und strategische Arbeit bleibt. Es wird angenommen, dass Menschen weiterhin für kritisches Denken und die Überprüfung von KI-Ergebnissen wichtig sind. Auch wird angenommen, dass neue technologische Anforderungen entstehen und Standardisierung die Kreativität einschränken kann.

Kernaussagen

KI in der Evaluation: Chance, Herausforderung, nächsten Schritte



2 Hauptarbeitsgruppen & Netzwerktätigkeit

Überblick über neue Arbeitsgruppen und Fachaustausch

1. Hauptgruppe: Allg. Grundsätze & Organisation

1. Erarbeitung einer Struktur und Inhalte für den Leitfaden
2. Organisation Jahreskongress – SEVAL
3. Beitrag beim Jahreskongress
4. KI in den SEVAL-Standards
5. Umgang mit KI in Organisationen

2. Hauptgruppe: Anwendungen von KI in der Evaluation, inkl. Use Cases, Best Practices

1. Vorstellung von definierten Tools und Konzepten (z. B. MAXQDA)
2. Überblick über Einsatz von KI in der Evaluation (u. a. Chancen)
3. Anwendungen von Tools in der qualitativen Forschung
4. Anwendungen von Tools in der quantitativen Forschung
5. Anwendung von Tools im Schreibprozess

Fazit: Zentrale Herausforderungen betreffen auch die Evaluationsstandards.

Generierung & Verbreitung von Wissen



Diskussionspunkte

1. Wie kann die Entwicklung und Vermittlung von technischen, methodischen und ethischen Kompetenzen für den verantwortungsvollen Einsatz von KI in der Evaluation gezielt unterstützt werden?
2. Welche weiteren Möglichkeiten bestehen, um den kontinuierlichen Austausch und Wissenstransfer zu praktischen KI-Anwendungen – insbesondere auch über ChatGPT hinaus – innerhalb der Evaluationspraxis zu fördern?
3. Welchen Einfluss hat der Einfluss von KI-Technologie auf die SEVAL-Standards?

Wir laden Evaluierende und Evaluationsverantwortliche aus der Bundesverwaltung herzlich ein, in den kontinuierlichen Austausch einzutreten – als Mitglied der AG, zum Erfahrungsaustausch und bei weiteren Anlässen!

Ausarbeitung in Zusammenarbeit mit der EEA





Disclaimer

Dieses Dokument wurde nach bestem Wissen und Gewissen der Unterzeichner erstellt. Soweit möglich, wurden öffentliche Quellen verwendet und als solche aufgeführt, Daten und Annahmen wurden entsprechend überprüft. Einige Informationen berufen auf private, nicht dokumentierten Aussagen von Marktteilnehmern und Experten, die entsprechend gekennzeichnet sind. Die AutorInnen übernehmen keine Haftung für die Vollständigkeit, Aktualität, Qualität und Richtigkeit solcher Aussagen.