



Universität St.Gallen

Institut für Bildungsmanagement
und Bildungstechnologien

Arbeiten mit smarten Assistenzsystemen

Ergebnisse einer 'Pulsmessung' und Erfahrungen
aus der Zusammenarbeit mit GPT-4 als Analyseassistent

Dr. Christoph Meier
St.Gallen, 17. April 2024



Inhaltsverzeichnis

Abbildungsverzeichnis	II
Management Summary	1
1 Einleitung	2
1.1 Generative KI im Arbeitsfeld	2
1.2 SCIL-Entwicklungspartnerschaft «Generative KI & Personalentwicklung»	2
1.3 Über diesen Arbeitsbericht	3
2 "Wie arbeitest du mit smarten Assistenzsystemen?" Eine 'Pulsmessung'	4
2.1 Instrument und Beteiligung	4
2.2 Ergebnisse	5
2.2.1 Bekanntheitsgrad & Nutzen	5
2.2.2 Eigenes Erleben bei der Nutzung der Werkzeuge	6
2.2.3 Sorgen	10
2.2.4 Erforderliche Unterstützung für produktive Nutzung	11
2.2.5 Themen der Kommentare	14
2.3 Zusammenfassung	15
3 Datenanalysen mit smarten Assistenzsystemen	17
3.1 Übersicht zum Vorgehen	17
3.2 Analyse von Daten zur Pulsmessung mit GPT-4	18
3.2.1 Vorbereitung	18
3.2.2 Konfiguration eines CustomGPT	18
3.2.3 Prompting in deutscher oder englischer Sprache?	19
3.2.4 Gestaltung der Prompts	20
3.2.5 Ergebnisvisualisierungen	21
3.2.6 Auswertung offener Kommentare und Sentiment Analyse	21
3.2.7 Überprüfung der Ergebnisse	23
3.3 ChatGPT&Co als Assistenten für Datenanalysen	25
4 Implikationen für Verantwortliche in Bildung und Personalentwicklung	26
Literaturverzeichnis	27
Anhang	29
Ausgewählte Kommentare zum Erleben der Zusammenarbeit mit generativer KI	29

Abbildungsverzeichnis

Abbildung 1: Verteilung der Beteiligung an der Pulsmessung.....	5
Abbildung 2: Häufigkeit der Nutzung von Typen von Werkzeugen auf Basis generativer KI	6
Abbildung 3: Wahrnehmungen der Nutzung von Werkzeugen auf Basis generativer KI	7
Abbildung 4: Erleben der Nutzung von generativer KI nach Beschäftigtengruppen.....	8
Abbildung 5: Erleben der Nutzung von generativer KI nach Altersgruppen	9
Abbildung 6: Häufigkeit der Aussage «mache mir (eher) Sorgen» nach Beschäftigtengruppe	10
Abbildung 7: Häufigkeit der Aussage «mache mir (eher) Sorgen» nach Beschäftigtengruppe	11
Abbildung 8: Erforderliche Unterstützung in der Arbeit mit generativer KI	12
Abbildung 9: Benötigte Unterstützung für produktive Nutzung generativer KI	13
Abbildung 10: Erforderliche Unterstützung nach Grad der Sorge bezüglich generativer KI	13
Abbildung 11: Inhaltsanalyse der Kommentare: Kategorien und Häufigkeit	14
Abbildung 12: Beispiele für Kommentare der Befragten	15
Abbildung 13: Vorgehen bei der Analyse der Daten zur Pulsmessung.....	18
Abbildung 14: Vom Prompt zum Python Code zur Ergebnisvisualisierung.....	19
Abbildung 15: Beispiele für frühe und fehlerhafte Visualisierungen zu Auswertungen.....	21
Abbildung 16: Sentiment-Analyse der offenen Kommentare zur Pulsmessung	23
Abbildung 17: Verteilung der Teilnehmenden nach Organisation, Berufsrolle und Alter (1)	24
Abbildung 18: Verteilung der Teilnehmenden nach Organisation, Berufsrolle und Alter (2)	24

Management Summary

Anwendungen generativer KI (ChatGPT&Co) verbreiten sich schnell im Arbeitsfeld (z.B. die verschiedenen Copiloten von Microsoft). Studien zeigen ein grosses Nutzenpotenzial im Hinblick auf Arbeitsproduktivität und ~qualität. Die kompetente Nutzung von bzw. Zusammenarbeit mit ChatGPT&CO («versierte Koaktivität») wird zu einer wichtigen Zukunftskompetenz.

Im Rahmen der SCIL-Entwicklungspartnerschaft 2023-24 "Generative KI & Personalentwicklung" wurde eine Befragung zur Nutzung generativer KI durch Mitarbeitende in den beteiligten Unternehmen durchgeführt. Beteiligt haben sich 1040 Personen in fünf Unternehmen aus unterschiedlichen Beschäftigtengruppen. Der vorliegende Bericht behandelt zwei Themen:

- (1) Ergebnisse der Befragung;
- (2) Erfahrungsbericht zum Einsatz von GPT-4 als Analyse-Assistent.

Zentrale Ergebnisse der Befragung sind folgende:

- (1) Von den verschiedenen Typen von Werkzeugen auf Basis generativer KI sind aktuell nur Text-Generatoren weithin bekannt.
- (2) Die Befragten erleben die Arbeit mit GenKI-Werkzeugen als (eher) spannend und sie erleben sich dabei zu grossen Teilen als (eher) produktiver.
- (3) Die Haltung zu Werkzeugen generativer KI und deren Nutzung ist neutral-sachlich bis leicht positiv. Deutlich negative Stimmen sind klar in der Minderheit.
- (4) Für eine produktive Nutzung von KI-Werkzeugen wünschen sich die Befragten vor allem Möglichkeiten, diese auszuprobieren, Beispiele guter Praxis und Gelegenheiten für Austausch mit Kolleg:innen.
- (5) Es zeigen sich Unterschiede zwischen verschiedenen Beschäftigten- und Altersgruppen.

Für die Datenanalysen und Ergebnisvisualisierungen wurde der Einsatz von GPT-4 als Assistenzsystem erprobt und reflektiert. Wichtige Schritte dabei waren die Konfiguration eines CustomGPT, das Prompting, die Ergebnisvisualisierung und schliesslich die Überprüfung der Ergebnisse. Es zeigt sich, dass GPT-4 hier umfangreiche Unterstützung bieten kann.

Die wichtigsten Implikationen für Verantwortliche in Bildung und Personalentwicklung sind:

- (1) Es braucht organisationsweite Initiativen zu Assistenzsystemen auf Basis generativer KI, zu Einsatzmöglichkeiten, Potenzialen und Limitationen.
- (2) Die unterschiedlichen Wahrnehmungen von Beschäftigtengruppen zu kollaborativem Arbeiten mit GenKI-Werkzeugen müssen bei der Konzeption und Umsetzung von entsprechenden Informations- und Befähigungsprogrammen berücksichtigt werden.
- (3) Verantwortliche für Bildung und Personalentwicklung sollten die Möglichkeiten, die Assistenzsysteme auf Basis generativer KI im Hinblick auf Datenanalysen und Ergebnisvisualisierungen bieten, aktiv erkunden und gegebenenfalls für sich selbst nutzen.

1 Einleitung

1.1 Generative KI im Arbeitsfeld

Generative künstliche Intelligenz markiert eine nächste Entwicklungsstufe von künstlicher Intelligenz. Als künstliche Intelligenz werden Systeme bezeichnet, welche intelligentes Verhalten zeigen, indem sie ihre Umwelt analysieren und Aktionen ausführen. Beispiele sind etwa autonom agierende Industrieroboter, selbstfahrende Autos oder Software für diagnostische Bildanalyse in der Medizin. Als generative KI werden Systeme bezeichnet, die nicht nur Daten analysieren, sondern auch neue Daten oder Inhalte hervorbringen können. Also beispielsweise Texte, Bilder, Präsentationen, Videos oder Computer Code (Seufert & Handschuh 2024).

Anwendungen generativer KI (ChatGPT&Co) verbreiten sich derzeit schnell im Arbeitsfeld (Kreacic et al. 2024) - unter anderem durch die Angebote von Microsoft wie copilot.microsoft.com und Copilot Pro. Diese Anwendungen basieren auf grossen Sprachmodellen sowie Transformer-Architekturen und werden auch als «smarte Assistenzsysteme» bezeichnet. ChatGPT ist vermutlich die bekannteste Anwendung dieser Klasse von Systemen. Diese Systeme sind "general purpose technologies" (Gmyrek et al., 2023), die für viele verschiedene Aufgaben und durch viele verschiedene Berufsgruppen genutzt werden können. Nicht zuletzt deshalb wird auch von einer fünften industriellen Revolution gesprochen (Cook et al. 2023).

Erste Studien zeigen einen grossen möglichen Nutzen im Hinblick auf Arbeitsproduktivität und Arbeitsqualität (Noy und Zhang 2023). Die kompetente Nutzung von bzw. Zusammenarbeit mit Assistenzsystemen auf Basis generativer KI – die SWK (2024) spricht diesbezüglich von «versierter Koaktivität» - wird als eine wichtige Zukunftskompetenz gesehen. Und diese Zukunftskompetenz zu fördern ist eine Aufgabe der Funktionsbereiche Bildung bzw. Personalentwicklung in Unternehmen und Organisationen.

1.2 SCIL-Entwicklungspartnerschaft «Generative KI & Personalentwicklung»

Vor diesem Hintergrund ist die SCIL-Entwicklungspartnerschaft 2023-24 «Generative KI & Personalentwicklung» im Spätherbst 2023 gestartet. Beteiligt sind Verantwortliche für Personalentwicklung und IT-Services der folgenden Unternehmen: AXA Schweiz, Baloise Group, Die Schweizerische Post, PostFinance, Schweizerische Bundesbahnen, Swisscom und Zurich Schweiz. Ziel der auf neun Monate angelegten gemeinsamen Arbeit ist es, das Themenfeld «Befähigung für erfolgreiche Zusammenarbeit mit smarten Assistenzsystemen» zu strukturieren und für priorisierte Herausforderungen Antworten zu erarbeiten bzw. Lösungen zu entwickeln. Beispiele für priorisierte Herausforderungen sind unter anderem die folgenden:

- die Klärung der Implikationen von generativer KI für die Personalplanung;
- die Exploration und Entwicklung von Use Cases zu generativer KI;

- die Sichtung und Sammlung von Einsatzszenarien für generative KI über den gesamten Employee Life Cycle;
- die Systematisierung von relevanten Lerninhalten zum Thema «Arbeiten mit generativer KI»;
- die Exploration des Potenzials von Anwendungen generativer KI für das Kuratieren von Lernressourcen;
- die Exploration von veränderten Aufgabenprofilen durch generative KI und den Implikationen für das Führen von Mitarbeitenden;
- die Exploration von Vorgehensweisen zur Förderung von AI Literacy;
- die Entwicklung und Erprobung von CustomGPTs für personalisiertes Tutoring und für personalisierte Skills Assessments.

1.3 Über diesen Arbeitsbericht

Eine der gemeinsamen Aktivitäten im Rahmen der SCIL-Entwicklungspartnerschaft 2023-24 war die Vorbereitung und Durchführung einer Pulsmessung zur Nutzung von generativer KI durch die Beschäftigten der beteiligten Unternehmen. Wir sprechen von einer 'Pulsmessung', weil das Ziel darin bestand, eine sehr einfache, kurze und schnell zu bearbeitende Umfrage zu entwickeln und durchzuführen. Dazu wurde ein Frageinstrument entwickelt, das von den Partnern übernommen und zum Teil noch im Detail angepasst wurde. Die Rückläufe wurden dann zusammengeführt, bereinigt und bildeten so die Grundlage für eine gesamthafte Auswertung. Die Ergebnisse dieser Auswertung machen den ersten Teil dieses Berichts aus. In einem zweiten Teil des Berichts folgt eine Darstellung und Reflexion des Vorgehens bei Datenanalyse und Ergebnisvisualisierungen – nämlich in Zusammenarbeit bzw. Koaktivität mit GPT-4.

2 "Wie arbeitest du mit smarten Assistenzsystemen?"

Eine 'Pulsmessung'

2.1 Instrument und Beteiligung

Die Befragung war bewusst als kurze, einfache und schnelle 'Pulsmessung' und nicht als wissenschaftliche Studie (im Sinne einer detaillierten Blutbild-Analyse) angelegt. Das Frage-Instrument wurde in MS Forms umgesetzt, weil dieses Format von allen Partner-Organisationen einfach kopiert und intern genutzt werden konnte. Die Befragung wurde von den Partnern im Zeitraum zwischen Januar und März 2024 jeweils über interne Kommunikationskanäle verfügbar gemacht. Beispielsweise über entsprechende Hinweise auf unternehmensinternen Intranet-Seiten oder über interne social Media-Kanäle. Die Beschäftigten wurden dabei zu folgenden Punkten befragt:

- Alter;
- berufliche Rolle;
- welche Typen von Assistenzsystemen auf Basis generativer KI bekannt sind bzw. genutzt werden;
- wie die Entwicklungen im Bereich generativer KI allgemein eingeschätzt werden;
- wie es erlebt wird, wenn mit Assistenzsystemen auf Basis generativer KI gearbeitet wird;
- was bzw. welche Art von Unterstützung es braucht, um mit Assistenzsystemen auf Basis generativer KI produktiv arbeiten zu können.

Bei der Pulsmessung handelt es sich nicht um eine Umfrage, die die Struktur der Belegschaften in den beteiligten Unternehmen repräsentativ abbildet. Vielmehr basieren die Rückläufe auf einer Selbstselektion der Teilnehmenden. Dies ist bei der Interpretation der Ergebnisse im Auge zu behalten. Der Median der Antwortzeiten liegt bei etwas mehr als dreieinhalb Minuten. Dies zeigt, dass es gelungen ist, eine kurze, sehr schnell zu bearbeitende Umfrage zu erstellen.

An der Pulsmessung haben sich insgesamt 1040 Personen beteiligt. Wie Abbildung 1 zeigt, verteilen sich die Rückläufe ungleichmässig auf

1. die fünf teilnehmenden Unternehmen (codiert als A, B, D, E, H);
2. drei grosse Beschäftigtengruppen, die unterschieden wurden
 - a) Verantwortung für administrative / operative Aufgaben ("Admin. / Operat.");
 - b) Verantwortung für Fach-Themen und / oder Projekte ("Proj. / Themen"); und
 - c) Verantwortung für Personal ("Führung").
3. unterschiedliche Altersgruppen.

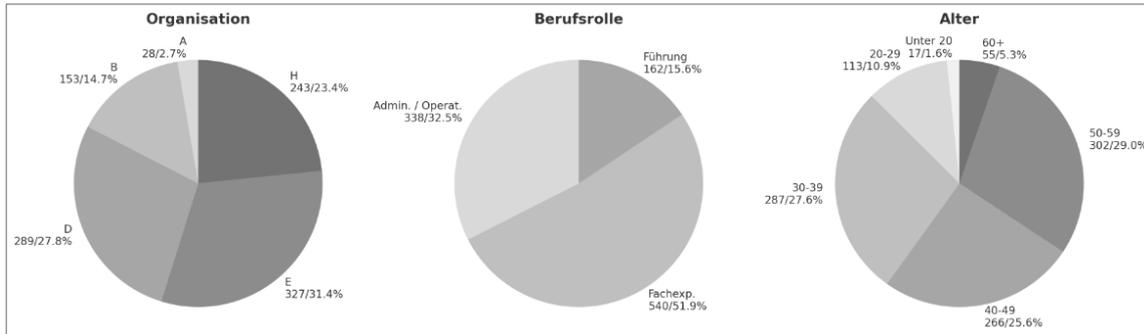


Abbildung 1: Verteilung der Beteiligung an der Pulsmessung (Bildquelle: SCIL)

2.2 Ergebnisse

2.2.1 Bekanntheitsgrad & Nutzen

Eine erste inhaltliche Frage der Pulsmessung betraf den Bekanntheitsgrad verschiedener Typen von GenKI-Anwendungen und die Intensität von deren Nutzung. Unterschieden wurden die folgenden Werkzeugtypen:

- "Text-Generatoren (ChatGPT, Bing, Claude, etc.)",
- "Bild-Generatoren (DALL-E, Midjourney, Stable Diffusion, etc.)",
- "Präsentations-Generatoren (TomeAI, BeautifulAI, etc.)",
- "Video-Generatoren (Synthesia, FlexiClip, etc.)",
- "Code-Generatoren (GitHubCoPilot, Code Whisperer, etc.)".

Die Ergebnisvisualisierung zeigt, dass die verfügbaren Typen von Werkzeugen auf Basis von GenKI grossenteils unbekannt sind. Abgesehen von Text-Generatoren (und teilweise auch von Bildgeneratoren) wurden diese Werkzeuge bisher nur von einer kleinen Minderheit der Befragten genutzt.

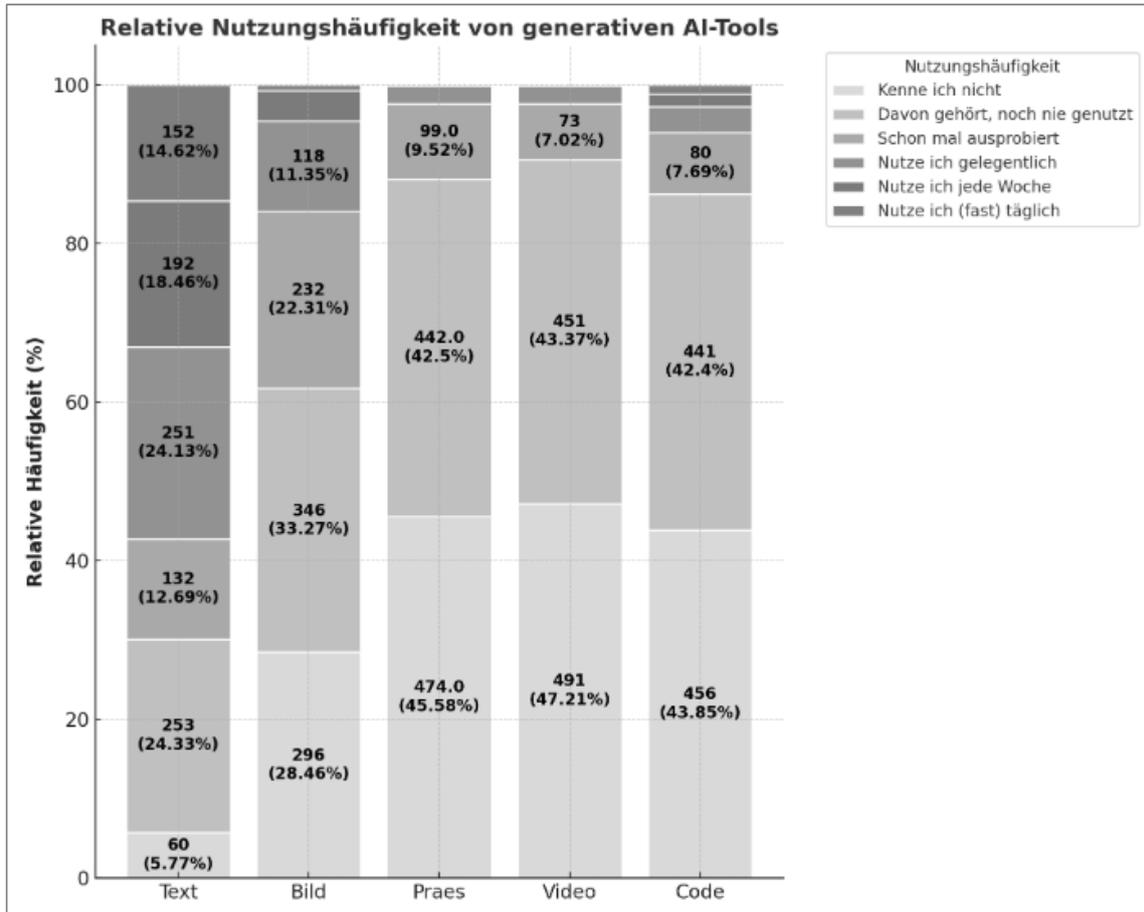


Abbildung 2: Häufigkeit der Nutzung von Typen von Werkzeugen auf Basis generativer KI (Bildquelle: SCIL)

Hinweis: Balken-Segmente, die weniger als 5% der Antworten in einer Kategorie repräsentieren, sind im Sinne einer besseren Lesbarkeit der Grafik nicht beschriftet.

Darüber hinaus zeigt eine weitere Auswertung, dass die Nutzungsintensität von Werkzeugen für Texterzeugung und Bilderzeugung bei den Beschäftigten in der Rolle «Verantwortung für administrative / operative Aufgaben» geringer ausfällt als bei den beiden anderen Rollen ("Proj. / Themen" und "Führung").

2.2.2 Eigenes Erleben bei der Nutzung der Werkzeuge

Eine zweite Frage betraf das Erleben der Beschäftigten, wenn sie Werkzeuge auf der Grundlage von generativer KI nutzen. Hier standen folgende Antwortoptionen zur Auswahl:

- "... fühle ich mich gestärkt (ich kann Aufgaben bewältigen, die ich sonst nicht schaffen könnte).";
- "... habe ich das Gefühl, produktiver zu sein.";
- "... finde ich das spannend (ich kann Neues entdecken und ausprobieren).";
- "... hätte ich gerne klare Vorgaben zum Was? und Wie?";
- "... komme ich mir vor, als ob ich schummeln würde.";
- "... frage ich mich, ob es mich und meine Arbeitskraft künftig noch braucht.".

Etwa 2/3 der Befragten erleben sich in der Nutzung von GenKI-Werkzeugen (eher) produktiver, gut 50% erleben sich als (eher) gestärkt. Weniger als 1/3 der Befragten erleben die Nutzung (eher) als ein Schummeln und sie fragen sich, ob sie entbehrlich werden.

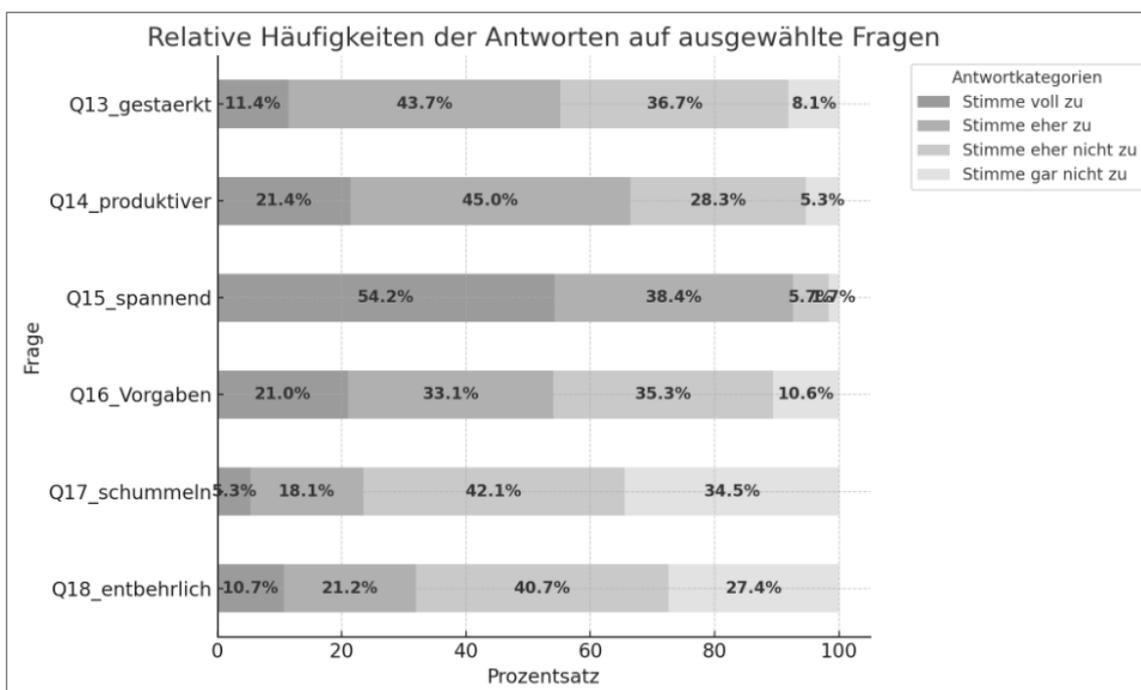


Abbildung 3: Wahrnehmungen der Nutzung von Werkzeugen auf Basis generativer KI (Bildquelle: SCIL)

Qualitative Studien in Unternehmen und Organisationen haben gezeigt, dass unterschiedliche Beschäftigtengruppen sehr unterschiedliche Perspektiven auf die Nutzung von (generativer) KI im Arbeitsfeld haben können. Beispielsweise, indem sie neue Geschäftsmöglichkeiten sehen, die Möglichkeit, (lästige) Routinearbeiten abzugeben, oder die Gefahr von lästiger Zusatzarbeit (Ei-nola & Khoreva, 2023).

Eine Auswertung zu den Wahrnehmungen bei der Nutzung unter Berücksichtigung der beruflichen Rollen der Befragten ergab Folgendes (vgl. Abb. 4):

- Mehr Zustimmung bei Fachexpert:innen / Projektverantwortlichen und Führungskräften zu den Selbstwahrnehmungen «fühle mich gestärkt» und «fühle mich produktiver»;
- Mehr Zustimmung bei Beschäftigten mit Verantwortung im Bereich administrativer bzw. operativer Aufgaben zu den Selbstwahrnehmungen «hätte gerne Vorgaben», «kommt mir vor wie schummeln» und «frage mich, ob es mich künftig noch braucht».

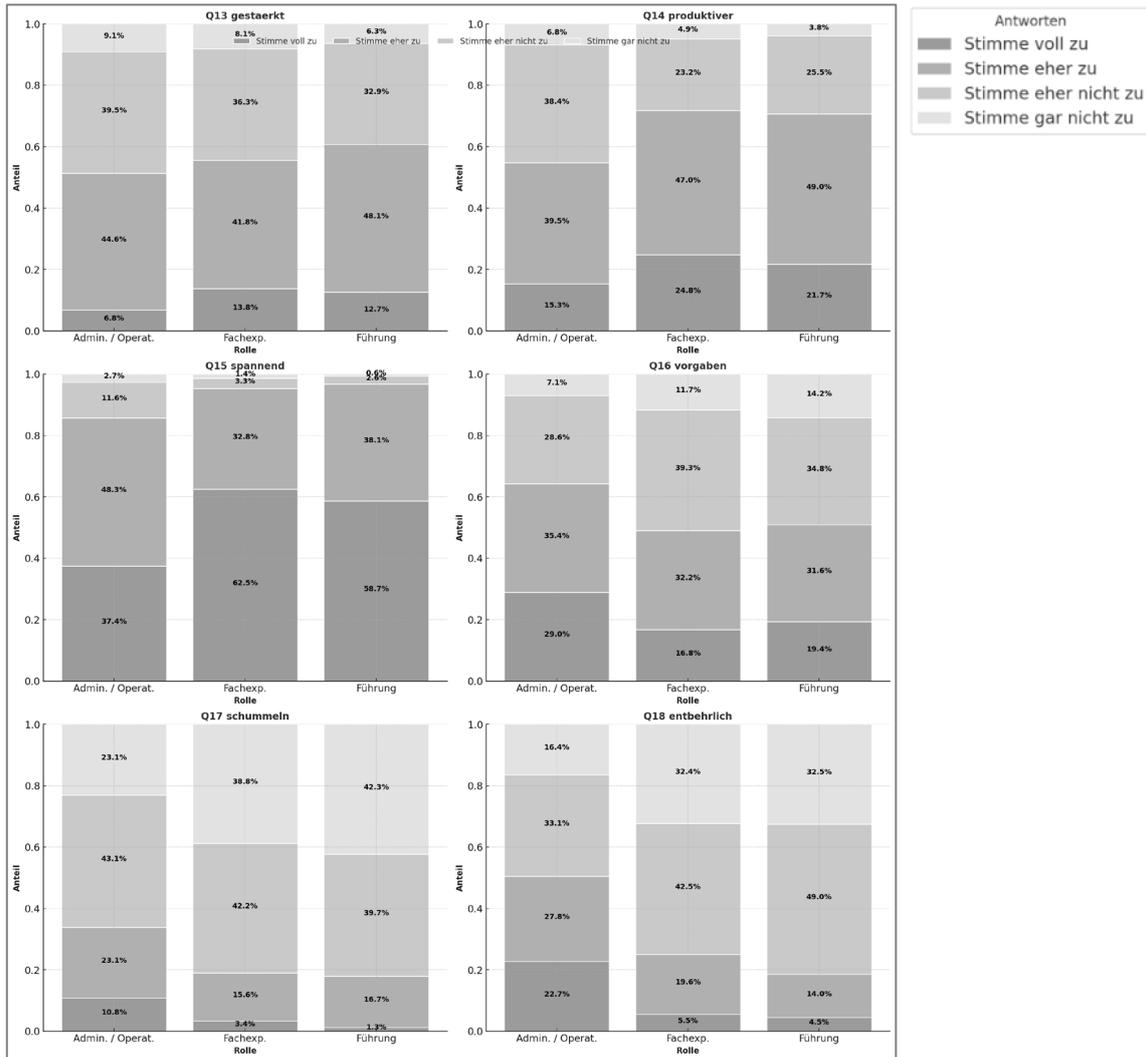


Abbildung 4: Erleben der Nutzung von generativer KI nach Beschäftigtengruppen (Bildquelle: SCIL)

Betrachtet man die Antworten auf diese Fragen getrennt nach Altersgruppen, dann fallen zwei Dinge auf (vgl. Abbildung 5):

- Die Zustimmung zur Aussage «fühle mich produktiver» ist für die Altersgruppe der 20-29-jährigen am höchsten und nimmt dann mit zunehmendem Alter kontinuierlich ab.
- Die Zustimmung zur Aussage «hätte ich gerne Vorgaben zum Was? und Wie?» nimmt mit zunehmendem Alter zu.

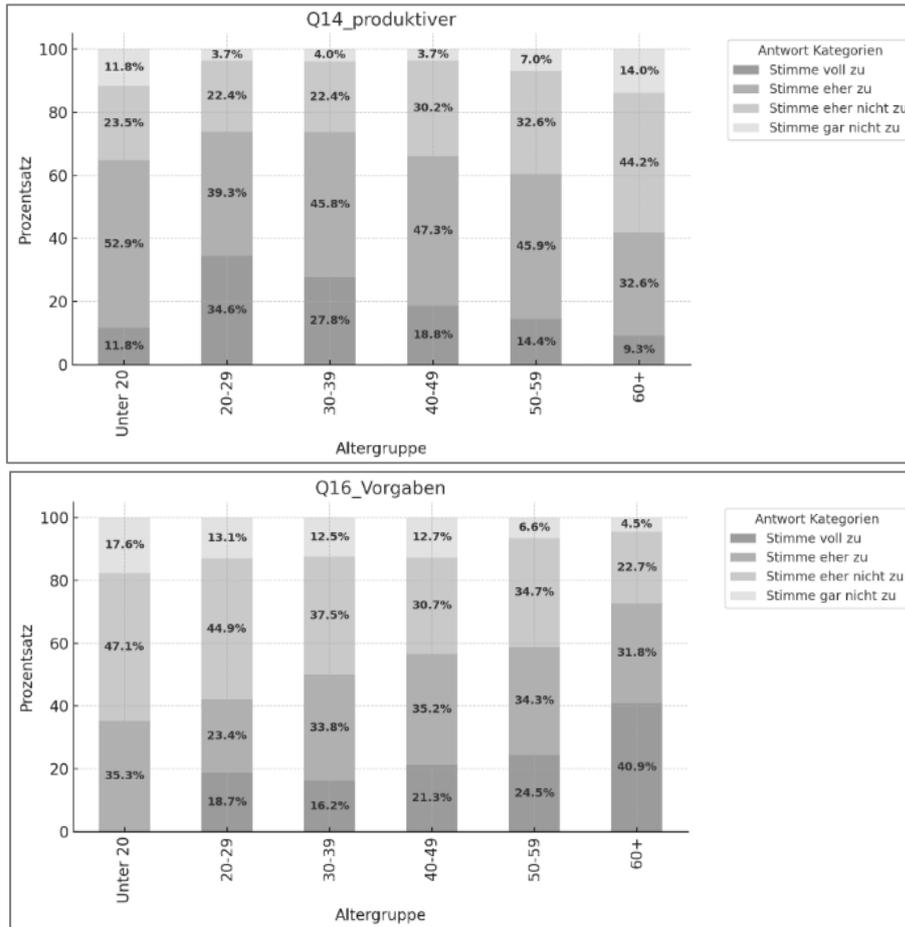


Abbildung 5: Erleben der Nutzung von generativer KI nach Altersgruppen (Bildquelle: SCIL)

Zum eigenen Erleben der Nutzung von generativer KI im Arbeitsfeld hatten wir auch eine offene Frage in die Pulsmessung aufgenommen.

Von den insgesamt 152 auswertbaren Antworten waren 118 in Deutsch, 23 in Französisch, 7 in Italienisch und 3 in Englisch gehalten. Hier jeweils drei Beispiele für positive und für negative Kommentare:

"Ein Werkzeug, dass man kennen sollte (speziell Prompt-Engineering), ohne gute Kenntnisse wird es bald schwierig, denke ich."

"Als Unterstützung möchte ich das gerne kennenlernen (zum Briefe schreiben, zum Kürzen von Texten...) aber IMMER vorbehaltlich einer guten Nachbearbeitung. Als Grundlage zum Weiterarbeiten super, aber die wenigen Texte, die ich gesehen habe, waren sehr wohl gut erkennbar als Produkt von KI."

"Es kommt immer ganz auf die Fragestellung an. So kann generative KI einerseits sehr schnell weiterhelfen und gute Inputs liefern, andererseits bleiben die erzeugten Antworten auch häufig unter den Erwartungen zurück und führen zu weiteren Abklärungen oder sind schlicht zu ungenau, als dass man sie verwenden könnte."

"Especially CoPilot sometimes provides wrong / useless suggestions. then it costs me more time because it interrupts the flow of coding. when I don't know about a solution I still have to look it up online to see what the problem was and why it works, if I want robust code and not just 'works at the moment, let's leave it'. the time gain is only if it adds closing brackets and stuff like that that is annoying to type."

"Ich finde, langfristig werden ganz viele Tätigkeitsfelder automatisiert werden können. [...] Es ist absehbar, dass immer weniger Tätigkeitsfelder übrigbleiben werden, die nicht automatisiert werden können. [...]"

"Non sono per tale 'progressi' perché sento i clienti raccontarmi episodi, esperienze"»
 ["I am not for such 'progress' because I hear customers tell me episodes, experiences"]
 ((Übersetzung durch DeepL.com))

2.2.3 Sorgen

Der Frage, wie die Beschäftigten die Entwicklungen im Bereich generative KI allgemein einschätzen, haben wir in der Pulsmessung ebenfalls angesprochen: "Wenn ich an die Entwicklung im Bereich der generativen künstlichen Intelligenz denke, dann ... ". Die beiden verfügbaren Antwortoptionen lauteten: "... mache ich mir Sorgen." und "... finde ich das spannend."

Die Analyse zeigt, dass sich aufgrund dieser Entwicklungen etwas mehr als 50% der Befragten in der Gruppe der Beschäftigten mit Verantwortung für administrative und operative Aufgaben (eher) Sorgen machen. Damit fällt der Anteil der sich sorgenden Personen deutlich höher aus als bei den beiden anderen Beschäftigtengruppen («Proj. / Themen» und «Führung»).

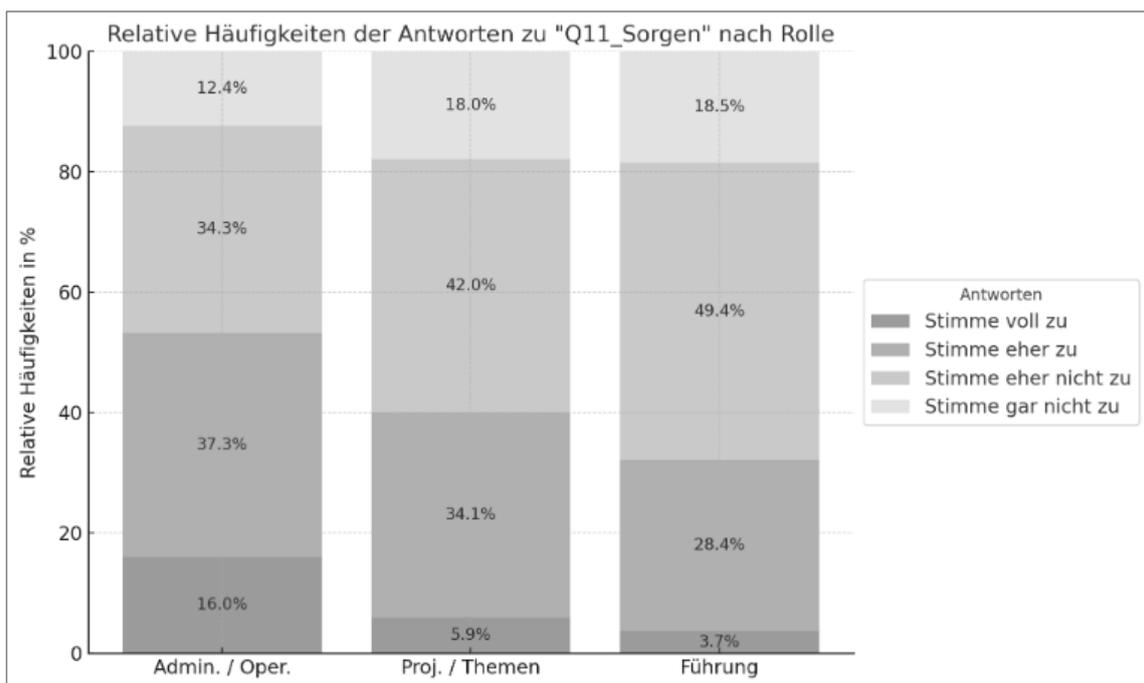


Abbildung 6: Häufigkeit der Aussage «mache mir (eher) Sorgen» nach Beschäftigtengruppe (Bildquelle: SCIL)

Weiter zeigt sich, dass ab der Altersgruppe der 40-jährigen der Anteil derjenigen, die sich aufgrund der Entwicklungen im Bereich generative KI (eher) Sorgen machen, kontinuierlich zunimmt.

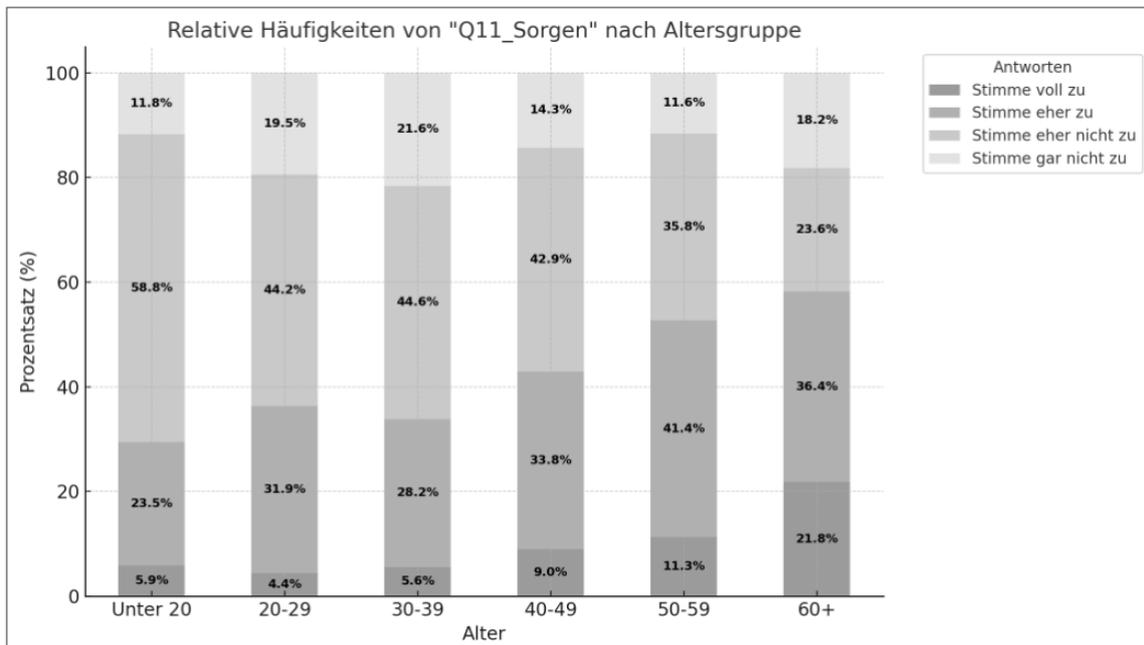


Abbildung 7: Häufigkeit der Aussage «mache mir (eher) Sorgen» nach Beschäftigtengruppe (Bildquelle: SCIL)

2.2.4 Erforderliche Unterstützung für produktive Nutzung

Eine letzte Auswahl-Frage der Pulsmessung betraf schliesslich die Art der Unterstützung, die es braucht, um produktiv mit Werkzeugen auf Basis generativer KI arbeiten zu können. Die Auswahl-Optionen waren die Folgenden:

- "... klare Vorgaben zum Was? und Wie?";
- "... passende digitale Lerninhalte (z.B. Video-Tutorials), die ich selbständig nutzen kann.";
- "... jemanden, der mir beibringt, wie ich diese Werkzeuge nutzen kann."; (in der Legende in Abb. 6 abgekürzt als "Coaching");
- "... Möglichkeiten, diese Werkzeuge in meinem Arbeitsalltag selbst auszuprobieren."
- "... Beispiele (gute Praxis), an denen ich mich orientieren kann.";
- "... Möglichkeiten, mich mit KollegInnen zur Nutzung dieser Werkzeuge auszutauschen."

Von den Antwortenden wurde den Optionen «brauche Möglichkeiten ... auszuprobieren», «brauche Beispiele guter Praxis» und «brauche Austausch mit Kolleg:innen» am häufigsten zugestimmt. Mit einigem Abstand folgen die Optionen "brauche passende Selbstlerninhalte", brauche klare Vorgaben" und "brauche jemanden, der mir beibringt, wie ich diese Werkzeuge nutzen kann." (in der Legende abgekürzt als «Coaching»).

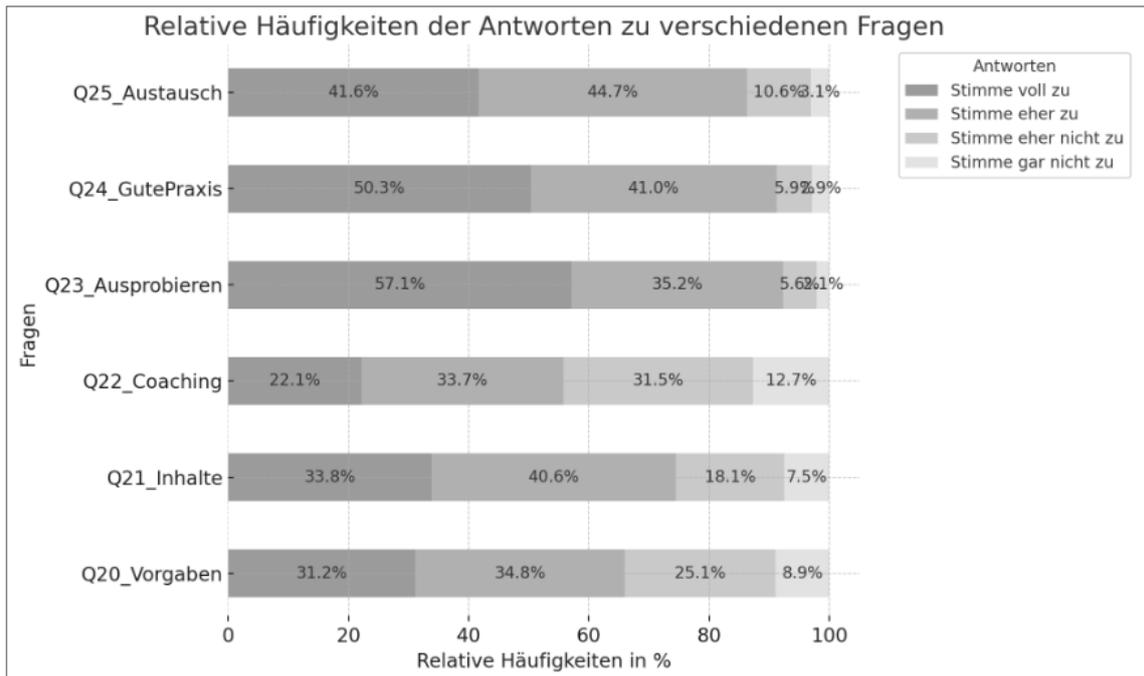


Abbildung 8: Erforderliche Unterstützung in der Arbeit mit generativer KI (Bildquelle: SCIL)

Ein Aspekt, der mit den Antwortoptionen nicht abgedeckt, aber in den offenen Kommentaren angesprochen wurde, ist die Integration in die technischen Systeme, die produktives Arbeiten unterstützen. Hier ein Auszug aus den Kommentaren dazu:

#219, Q26

«Was aus meiner Sicht am Meisten Fehlt, ist die Integration in unsere Systeme. Bsp. KI Zugriff auf all unsere Wissens- und Informationsablagen geben. Anschliessend das Beantworten von Fragen, generieren von Präsentationen, etc. zu ermöglichen. Dies wäre eine extreme Erleichterung.»

Eine Auswertung zur erforderlichen Unterstützung nach Altersgruppen zeigt folgendes:

- Die Zustimmung zu den genannten Unterstützungsangeboten nimmt mit zunehmendem Alter zu.
- Eine Ausnahme bildet die Gruppe der Beschäftigten in der Altersgruppe 60+. Hier zeigt sich im Hinblick auf manche Unterstützungsangebote („Lerninhalte zur selbständigen Bearbeitung“; „Möglichkeiten zum Ausprobieren“; „Beispiele guter Praxis“) eine Stagnation bzw. ein niedrigerer Wert als bei der vorherigen Altersgruppe.

Eine mögliche Interpretation für die zweite Beobachtung ist, dass ältere Beschäftigte das Thema «produktives Arbeiten mit generativer KI» für sich selbst als nur noch begrenzt relevant sehen. Abbildung 9 illustriert diese beiden Beobachtungen anhand von zwei Auswertungen:

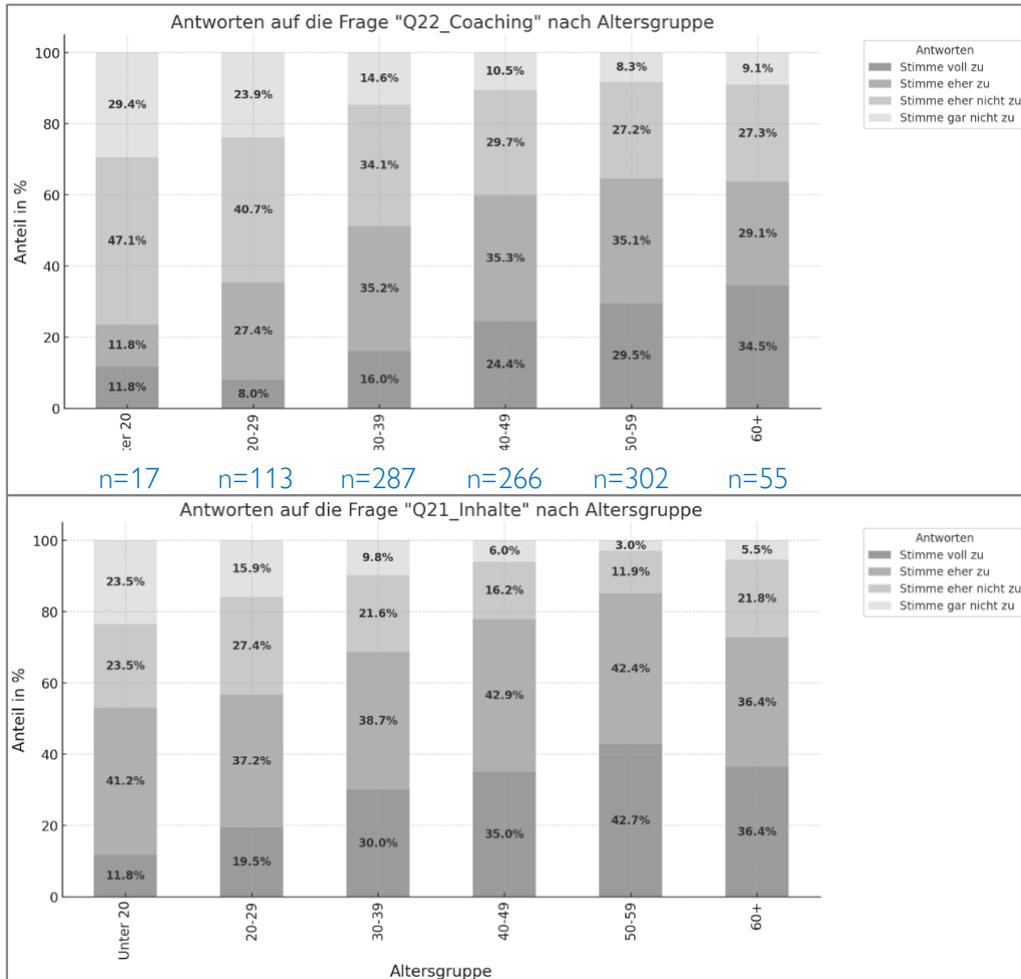


Abbildung 9: Benötigte Unterstützung für produktive Nutzung generativer KI (Bildquelle: SCIL)

Weitere Analysen zeigen, dass Beschäftigte, die sich aufgrund der Entwicklungen im Bereich generative KI Sorgen machen, Unterstützungserfordernisse anders priorisieren als diejenigen, die sich diesbezüglich keine Sorgen machen. Wie Abbildung 7 zeigt, wünschen sich Beschäftigte, die sich Sorgen machen (schraffierte Balken), deutlich häufiger "klare Vorgaben", "passende digitale Lerninhalte" und "jemanden, der mir beibringt ... (Coaching)":

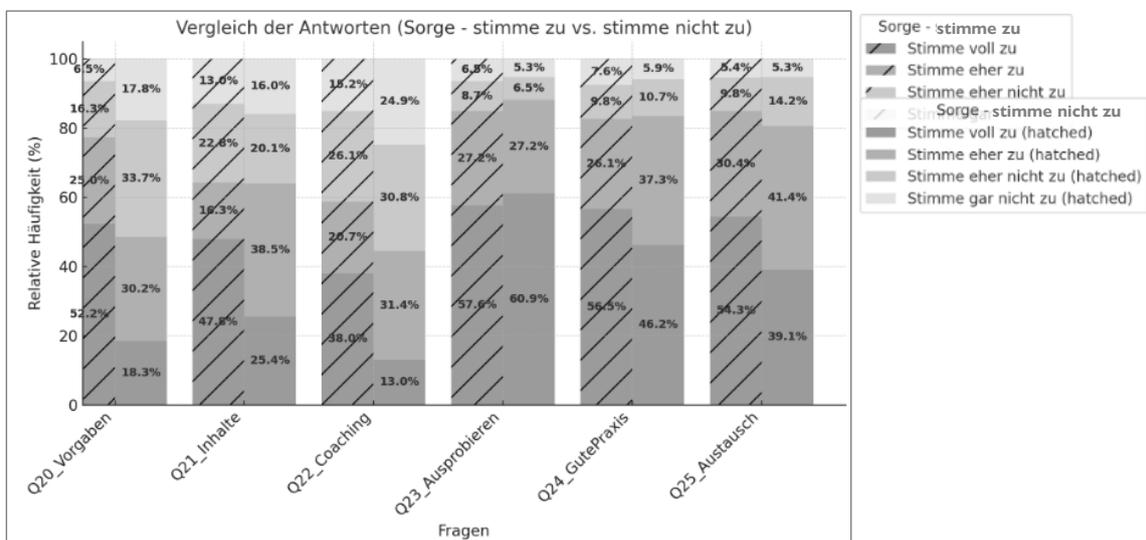


Abbildung 10: Erforderliche Unterstützung nach Grad der Sorge bezüglich generativer KI (Bildquelle: SCIL)

2.2.5 Themen der Kommentare

Im Rahmen der Pulsmessung wurden von den 1040 befragten Personen zahlreiche Kommentare in den Sprachen Deutsch, Französisch, Italienisch und Englisch abgegeben. In die Auswertung einbezogen wurden 161 Kommentare, die die Befragten in zwei Kommentarfelder eingegeben hatten (Kommentare zur Frage, wie sie das Arbeiten mit Werkzeugen generativer KI erleben; offene Kommentare ganz am Ende der Pulsmessung).

Um diese Kommentare analysieren zu können, wurden diese zunächst in eine separate Datei ausgelesen und mit Hilfe von DeepL.com ins Englische übersetzt. Eine schnelle und einfache erste Inhaltsanalyse ergab folgende Übersicht zu den Themen dieser Kommentare:

Category	Count
Perceived Benefits of AI	15
Perceived Risks and Concerns	20
AI Integration and Usage	18
Ethical and Legal Considerations	12
Ambivalence towards AI	10

Abbildung 11: Inhaltsanalyse der Kommentare: Kategorien und Häufigkeit (Bildquelle: SCIL)

Die nachfolgende Abbildung zeigt jeweils drei Beispiele für Kommentare der Befragten zu diesen Themen:

<p>1. Perceived Benefits of AI</p> <ul style="list-style-type: none"> • Count: 15 • Examples: <ul style="list-style-type: none"> • (19_ID 42): Advocacy for being a pioneer in using AI tools. • (19_ID 95): Empowerment due to efficient task management. • (19_ID 526): AI as a time-saving tool in the workplace. <p>2. Perceived Risks and Concerns</p> <ul style="list-style-type: none"> • Count: 20 • Examples: <ul style="list-style-type: none"> • (19_ID 50): Concerns about losing mental development. • (19_ID 163): Skepticism about AI as a black box. • (19_ID 687): Concerns about knowledge becoming "flatter" due to AI. <p>3. AI Integration and Usage</p> <ul style="list-style-type: none"> • Count: 18 • Examples: <ul style="list-style-type: none"> • (19_ID 108): Need for introduction to AI tools. • (19_ID 610): Learning about risks and opportunities of AI. • (27_ID 323): Suggestion for a "CHATConfluence". <p>4. Ethical and Legal Considerations</p> <ul style="list-style-type: none"> • Count: 12 • Examples: <ul style="list-style-type: none"> • (19_ID 146): Importance of regulations. • (27_ID 172): Importance of adhering to codes of ethics. • (27_ID 973): Fear about the direction of AI development. <p>5. Ambivalence towards AI</p> <ul style="list-style-type: none"> • Count: 10 • Examples: <ul style="list-style-type: none"> • (19_ID 97): AI as helpful but requires critical input. • (19_ID 474): Positive about AI but unsure about job security. • (27_ID 1028): AI seen as an exciting toy but also overhyped.

Abbildung 12: Beispiele für Kommentare der Befragten (Bildquelle: SCIL)

2.3 Zusammenfassung

Für den deutschsprachigen Raum liegen bislang kaum empirische Studien zur Nutzung von ChatGPT&Co im betrieblichen Kontext vor. Die gemeinsam im Rahmen der SCIL Entwicklungspartnerschaft 2023-24 durchgeführte Erhebung erhebt nicht den Anspruch einer Repräsentativität im Hinblick auf die Beschäftigten in den teilnehmenden Unternehmen. Vielmehr stellt sie eine Pulsmessung dar, die einfach gehalten und schnell zu wiederholen ist. Kernpunkte der vorliegenden Pulsmessung und der Auswertung sind die folgenden:

- (1) Beteiligt haben sich 1040 Personen aus fünf Unternehmen.
- (2) Die Befragten können unterschiedlichen Berufsgruppen zugeordnet werden:
 - Beschäftigte mit eher operativen / administrativen Aufgaben: ca. 33%;
 - Beschäftigte mit Aufgaben eher im Bereich Fachverantwortung / Projekte: ca. 50%;
 - Beschäftigte mit Personalverantwortung / Führungsaufgaben: ca. 15%.
- (3) Von den verschiedenen Typen von GenKI-Werkzeugen sind aktuell nur Text-Generatoren (und mit Einschränkungen auch Bildgeneratoren) den Befragten weithin bekannt.

- (4) Die Befragten erleben die Arbeit mit GenKI-Werkzeugen fast durchgängig als (eher) spannend und sie erleben sich dabei zu grossen Teilen als (eher) produktiver. Weniger als ein Drittel der Befragten fragt sich dabei, ob sie entbehrlich werden. Und weniger als ein Viertel der Befragten erleben die Nutzung (eher) als ein Schummeln.
- (5) Die Haltung zu GenKI-Werkzeugen und deren Nutzung, die aus den Antworten und auch aus den mitgelieferten Kommentaren hervorscheint, ist neutral-sachlich bis leicht positiv. Deutlich negative Stimmen sind klar in der Minderheit.
- (6) Im Hinblick auf die Frage, welche Unterstützung sie für eine produktive Nutzung von Werkzeugen generativer KI brauchen, zeigt sich folgendes Bild:
- die Möglichkeit, diese Werkzeuge auszuprobieren, steht klar im Vordergrund;
 - es folgen Beispiele guter Praxis und Austausch mit Kolleg:innen;
 - Vorgaben, Lerninhalte und Coaching wurden deutlich weniger oft genannt.
- (7) Es zeigen sich Unterschiede zwischen verschiedenen Beschäftigten- und Altersgruppen:
- a. Die Aussagen zum Erleben der Arbeit mit GenKI-Werkzeugen «fühle mich gestärkt» und «fühle mich produktiver» finden bei Fachexpert:innen / Projektverantwortlichen und Führungskräften mehr Zustimmung als bei den Beschäftigten mit eher operativen / administrativen Aufgaben.
 - b. Die Zustimmung zur Aussage «fühle mich produktiver» ist für die Altersgruppe der 20-29-jährigen am höchsten und nimmt dann mit zunehmendem Alter kontinuierlich ab.
 - c. Die Zustimmung zur Aussage «hätte ich gerne Vorgaben zum Was? und Wie?» nimmt mit zunehmendem Alter zu.
 - d. Ab der Altersgruppe der 40-jährigen nimmt der Anteil derjenigen, die sich aufgrund der Entwicklungen im Bereich generative KI (eher) Sorgen machen, kontinuierlich zu.
 - e. Beschäftigte mit Verantwortung für administrative und operative Aufgaben sind aufgrund der Entwicklungen im Bereich generative KI häufiger und stärker besorgt als andere Beschäftigtengruppen. Etwas mehr als 50% der Befragten in dieser Beschäftigtengruppe machen sind (eher) besorgt.
 - f. Beschäftigte, die sich aufgrund der Entwicklungen im Bereich der generativen KI Sorgen machen (Antwort «Stimme voll zu»), priorisieren Unterstützungserfordernisse anders als diejenigen, die sich diesbezüglich keine Sorgen machen. Erstere wünschen sich deutlich häufiger "klare Vorgaben", "passende digitale Lerninhalte" und "jemanden, der [es] mir beibringt".
 - g. Die Zustimmung zu den abgefragten Unterstützungsangeboten nimmt mit zunehmendem Alter zu.
 - h. Eine Ausnahme bildet die Gruppe der Beschäftigten in der Altersgruppe 60+. Hier zeigt sich im Hinblick auf manche Unterstützungsangebote („Lerninhalte zur selbständigen Bearbeitung“; „Möglichkeiten zum Ausprobieren“; „Beispiele guter Praxis“) eine Stagnation bzw. ein niedrigerer Wert als bei der vorherigen Altersgruppe.

3 Datenanalysen mit smarten Assistenzsystemen

Eines der Ziele der SCIL Entwicklungspartnerschaft 2023-24 besteht darin, die neu verfügbaren Werkzeuge auf Basis generativer KI soweit wie möglich auch für die eigenen Arbeiten zu erproben und zu nutzen. Im Rahmen der gemeinsamen Arbeit wurden diese Werkzeuge bisher beispielsweise für das Zusammenfassen von Studien zur Nutzung von generativer KI erprobt, für das Systematisieren von Use Cases für Anwendungen generativer KI, für das Pilotieren von personalisiertem Tutoring oder auch für das Kuratieren von Lerninhalten.

Die Analyse von Daten und die Visualisierung von Ergebnissen ist ein weiterer möglicher Anwendungsfall für GenKI-Werkzeuge, der im Funktionsbereich PE / L&D relevant ist. Natürlich gibt es dafür schon andere, zum Teil gut etablierte Werkzeuge (MS Excel, SPSS, MS PowerBI, etc.). Aber nicht alle Beschäftigten im Funktionsbereich PE / L&D sind im Umgang mit diesen Werkzeugen versiert (dazu später mehr). Darüber hinaus gibt es aktuelle Forschungsarbeiten, die der Frage nachgehen, inwiefern GPT-4 als Datenanalyst Menschen bereits ebenbürtig ist oder nicht (Cheng et al. 2023).¹

Im Rahmen der durchgeführten Pulsmessung wurde daher exploriert, inwiefern die erforderlichen Arbeitsschritte von GPT-4 unterstützt werden können. In den nachfolgenden Abschnitten ist dargestellt, wie bei der Datenanalyse vorgegangen wurde und welche Erfahrungen dabei gemacht wurden. Diese Darstellung ist als Momentaufnahme zu werten. Mit der Verbreitung von Microsofts Copilot Pro (beispielsweise in der Nutzung von MS Excel) und mit der Verfügbarkeit neuerer, leistungsfähigerer Sprachmodelle bzw. GPTs wird sich dieses Bild wieder ändern.

3.1 Übersicht zum Vorgehen

Die folgende Abbildung zeigt in einer Übersicht das Vorgehen bei der Analyse der Daten zur Pulsmessung:

¹¹ Cheng et al. kommen zu dem Schluss, dass es noch zu früh ist, diese Frage abschliessend zu beantworten. Sie formulieren folgendes Fazit zu ihrer Studie: «Experimental results show that GPT-4 can beat an entry-level data analyst and an intern data analyst in terms of performance and have comparable performance to a senior-level data analyst. In terms of the cost and time of our experiments, GPT-4 is much cheaper and faster than hiring a data analyst.» Sie weisen aber auch darauf hin, dass es bei GPT-4 zu Halluzinationen kommen kann und dass Berechnungen ungenau sein können (Cheng et al. 2023, S. 1, 12).

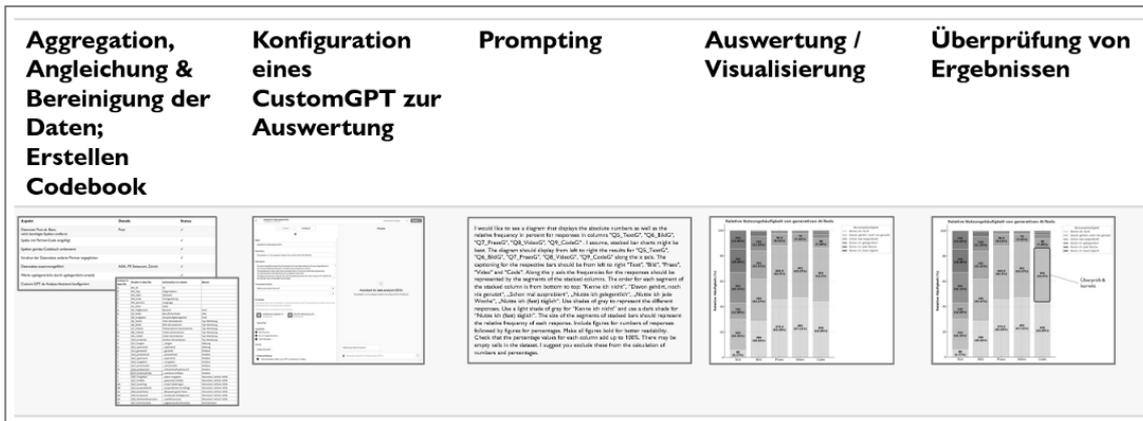


Abbildung 13: Vorgehen bei der Analyse der Daten zur Pulsmessung (Bildquelle: SCIL)

In den nächsten Abschnitten werden verschiedene Schritte erläutert und die dabei gemachten Erfahrungen festgehalten.

3.2 Analyse von Daten zur Pulsmessung mit GPT-4

3.2.1 Vorbereitung

Die vorbereitenden Aufgaben (u.a. das Zusammenführen und Angleichen der Datensätze der teilnehmenden Partner oder das Erstellen eines Codebook) wurden ohne Unterstützung durch GPT-4 umgesetzt. Inwiefern auch hierbei eine Unterstützung möglich gewesen wäre, wurde nicht exploriert.

3.2.2 Konfiguration eines CustomGPT

Für die ersten Auswertungen zur Pulsmessung wurde der Standard-Dialog in GPT-4 verwendet. Dabei wurde schnell klar, dass ein wiederholtes Einlesen der Datei mit den Umfragedaten zu viel Zeit in Anspruch nimmt. Daher wurde für die Arbeit mit diesem Datensatz ein CustomGPT konfiguriert ("Assistant for data analysis 02 En"), in dem der Datensatz, das zugehörige Codebook und grundlegende Anweisungen hinterlegt wurden.

Hier ein Auszug aus der Konfiguration des verwendeten CustomGPT:

«You are an assistant supporting the analysis of survey data relating to the use of generative AI. You conduct statistical analysis in a careful and conscientious manner. The analyses are mostly descriptive; perhaps there will be some inferential analyses also. You have access to the data file and to a codebook document. The data can be found in the file "[Name of file]" and on the sheet "Survey Data". You explicate results and generate diagrams that are easy to understand.

You observe the following specifics when generating diagrams:

Captions:

- (1) Make sure that all captions are in German language.
- (2) Always use larger font for all captions

(3) Abbreviate the professional roles as follows:

...»

Grundsätzlich erscheint es sinnvoll, Informationen, die über verschiedene Auswertungen hinweg gelten (z.B. zu den Farbtönen, die in den Diagrammen verwendet werden sollen), in die Konfiguration des CustomGPT aufzunehmen. Allerdings war zu Beginn der analytischen Arbeit am vorliegenden Datensatz nicht immer klar, welche Informationen es öfter braucht und welche nicht. Dies führte dazu, dass die Konfiguration des CustomGPT mehrfach angepasst wurde. Die bisherigen Erfahrungen zeigen, dass es eine Weile braucht, bis eine stabile Routine für die Aufteilung von Anweisungen zwischen CustomGPT und GPT-4-Prompt entwickelt wird.

3.2.3 Prompting in deutscher oder englischer Sprache?

Im Juli 2023 hat OpenAI eine Version von GPT verfügbar gemacht: GPT-4 Code Interpreter. Mittlerweile ist GPT-4 Code Interpreter integrierter Bestandteil des kostenpflichtigen Services GPT-4 oder GPTplus von OpenAI. Mit GPT-4 Code Interpreter ist es unter anderem möglich, Dateien als Input zu nutzen, Python Code auszuführen und Datenanalysen durchzuführen (Yosifova, 2024).

Der Weg zur Auswertung von Daten und zu Ergebnisvisualisierungen führt also über die folgenden Schritte (vgl. Abbildung 14):

- Formulierung eines Prompts für die gewünschte Auswertung bzw. Visualisierung
- Umsetzung des Prompts in Python Code durch GPT-4
- Ausführen des Python Codes in GPT-4
- Erzeugen von Ergebnisvisualisierungen durch GPT-4.

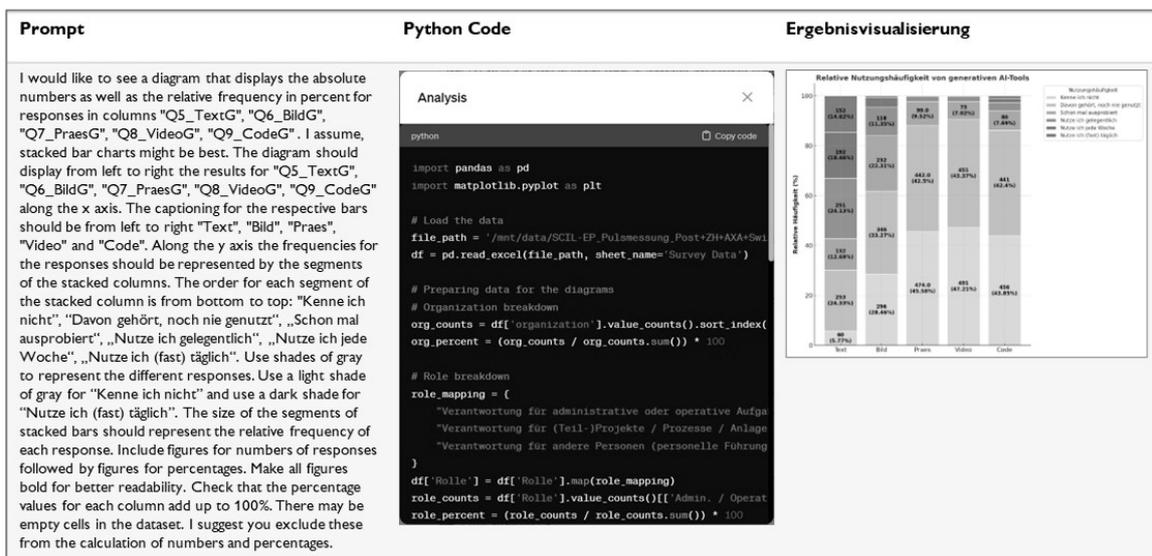


Abbildung 14: Vom Prompt zum Python Code zur Ergebnisvisualisierung (Bildquelle: SCIL)

Die Tatsache, dass eine Übersetzung von Prompts in natürlicher Sprache in Computer Code überhaupt möglich ist, ist eine der zentralen Errungenschaften der grossen Sprachmodelle. Die Konsequenzen sind weitreichend. Jensen Huang, der COE von Nvidia und andere haben dies pointiert formuliert:

“Human language is now computer programming code”

und

“Everyone is a programmer”(Jensen Huang)

“The hottest new programming language is English”

(Karpathy in deepakvraghavan, 2023).

Für die Qualität der Analysen und Ergebnisvisualisierungen ist folglich zentral, ob die im Prompt repräsentierte Intention in einer guten Weise in Python Code umgesetzt wird. Es ist daher wichtig, die Prompts so zu formulieren, dass die Intention bzw. Zielsetzung richtig erkannt wird.

Im Laufe der Exploration zur Datenauswertung mit GPT-4 wurde auch eine deutsche Fassung des CustomGPT erstellt und genutzt. Allerdings stellte sich schnell der Eindruck ein, dass die Analysen bzw. Ergebnisvisualisierungen mehr Fehler und Ungereimtheiten aufweisen (z.B. bei der Positionierung und Strukturierung von Legenden). Dies passt zusammen mit Beobachtungen, dass GPT-4 Englisch besser versteht als Deutsch. Die oben gezeigten Analysen und Auswertungen wurden daher alle mit der englischsprachigen Version des CustomGPT und mit englischsprachigen Prompts erzeugt.

3.2.4 Gestaltung der Prompts

Das Vorgehen beim Anstossen spezifischer Auswertungen (Prompting) hat sich im Lauf der Arbeit an der Auswertung der Pulsmessung verändert. Zu Beginn wurden teilweise spezifische Visualisierungen (z.B. Boxplots) eingefordert. Die Ergebnisse waren aber häufig nicht zufriedenstellend, so dass allgemeiner nach einer Auswertung bzw. Visualisierung gefragt wurde. Dabei haben sich gestapelte Balkendiagramme als gut lesbare Visualisierung herausgestellt und diese wurden in den Prompts wiederholt explizit angefragt. Hier der Prompt, der zu Abbildung 4 (oben) geführt hat:

«Create a diagram showing the relative frequencies of the answers to the questions "Q13_gestaerkt", "Q14_produkativer", "Q15_spannend", "Q16_Vorgaben", "Q17_schummeln" and "Q18_entbehrlich" by role according to «Q3_Rolle». For each role, create stacked bar charts showing the percentage of answers to "Stimme voll zu", "Stimme eher zu", "Stimme eher nicht zu" und "Stimme gar nicht zu". Create stacked bars for the following categories in this order: "Q13_gestaerkt", "Q14_produkativer", "Q15_spannend", "Q16_Vorgaben", "Q17_schummeln" und "Q18_entbehrlich". The colours in the bar charts represent the different response categories in this order: "Stimme voll zu" in dark grey (#9B9B9B), "Stimme eher zu" in medium grey (#B0B0B0), "Stimme eher nicht zu" in light grey (#C9C9C9) and "Stimme gar nicht zu" in very light grey (#E2E2E2).»

Es hat sich auch gezeigt, dass ein sehr sorgfältiges Formulieren wichtig ist und die Arbeiten beschleunigt. Tippfehler oder uneinheitliche Benennungen und Schreibweisen (z.B. "stimme voll zu" an Stelle von "Stimme voll zu") können dazu führen, dass Berechnungen abgebrochen und neu gestartet werden oder gar nicht erfolgreich durchgeführt werden können.

3.2.5 Ergebnisvisualisierungen

Die von GPT-4 erzeugten Ergebnisvisualisierungen waren nicht von Anfang an brauchbar, wie die Beispiele in Abbildung 15 zeigen:

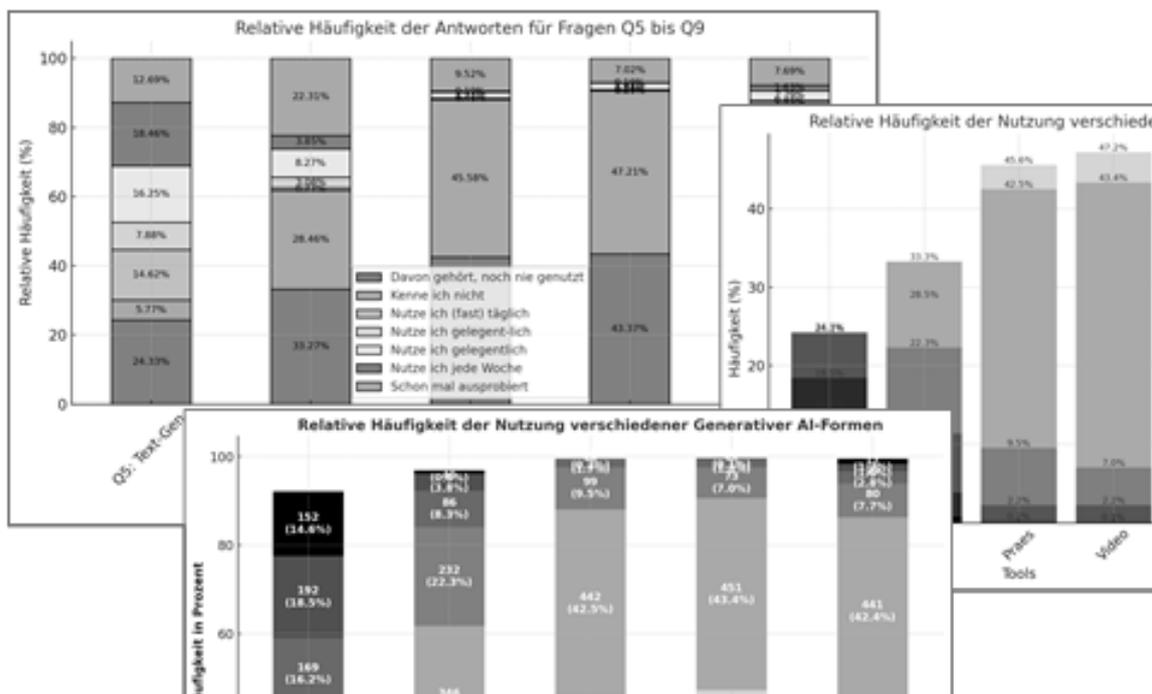


Abbildung 15: Beispiele für frühe und fehlerhafte Visualisierungen zu Auswertungen (Bildquelle: SCIL)

Mit zunehmender Dauer der Arbeit an den Auswertungen wurde deutlich, dass ein grosser Teil des Aufwands in der Arbeit mit GPT-4 darin besteht, den Prompt so zu verfeinern, dass die Ergebnisvisualisierungen in gut lesbarer Form ausgegeben werden. Beispielsweise im Hinblick auf die Abstufung von Farben oder die Verwendung von fett formatierten Schriften.

3.2.6 Auswertung offener Kommentare und Sentiment Analyse

Werkzeuge wie ChatGPT&Co können unter anderem auch für qualitative Datenanalysen eingesetzt werden (Morgan 2023) sowie für Inhaltsanalysen oder Sentiment Analysen – beispielsweise im digitalen Marketing (Krugmann & Hartmann 2024). Auf der einen Seite lockt die Aussicht auf Erleichterungen bei sehr aufwändigen Analysearbeiten. Auf der anderen Seite stehen Fragen nach Qualität, Reproduzierbarkeit und und Biases in maschinell erstellten Analysen (Morgan 2023).

Zhang et al. (2023) kommen zum Schluss, dass leistungsfähige LLMs bzw. GPTs auf der Grundlage gut gestalteter Prompts durchaus für qualitative Datenanalysen genutzt werden können. Und Krugmann und Hartmann (2024, S. 8) ziehen ein ähnliches Fazit im Hinblick auf Sentiment Analysen: «In sum, all three tested LLMs exhibit remarkable zero-shot sentiment classification accuracy in binary and three-class settings.» Sie weisen aber auch darauf hin, dass die Zuverlässigkeit von Sentiment Analysen bei der Analyse von kurzen Inhaltselementen niedriger ist als bei der Analyse von längeren Textpassagen: «Content-laden text documents, containing longer words and comprising multisentence documents, significantly increased sentiment prediction accuracy across LLMs, highlighting their proficiency with detailed and context-rich content.» (Krugmann & Hartmann 2024, S. 15).

Erste Versuche, die Kommentare der Teilnehmenden an der Pulsmessung zum Erleben der Arbeit mit GenKI-Werkzeugen in Kategorien einordnen zu lassen und Beispiele für jede Kategorie liefern zu lassen (vgl. den nachfolgenden Prompt), lieferten keine überzeugenden Ergebnisse. Mögliche Gründe sind die Kürze der Kommentare (häufig nur ein Satz) und die Tatsache, dass hier Antworten aus verschiedenen Sprachen (Deutsch, Französisch, Italienisch und Englisch) verarbeitet werden mussten.

Ausgehend von dieser Erfahrung wurde zunächst ein sprachlich homogenes Datenset für die Auswertung der offenen Kommentare erstellt:

- In einem ersten Schritt wurden mit dem GPT-4 Auswertungs-Assistenten zwei Textdateien erstellt, die jeweils alle Antworten zu zwei ausgewählten Kommentar-Feldern in der Pulsmessung sortiert nach Sprache zusammenführen und mit vorangestellten ID-Nummern versehen. Da die Sortierung der Kommentare nach den Sprachen fehlerbehaftet war, musste manuell nachgearbeitet werden.
- In einem zweiten Schritt wurden die Kommentare unter Verwendung von DeepL.com ins Englische übersetzt und manuell in einer einzigen Datei zusammengeführt.
- In einem dritten Schritt wurde die konsolidierte und übersetzte Datei für weitere Analysen in den Auswertungs-Assistenten hochgeladen.

Ein Beispiel für eine auf dieser Grundlage basierende Inhaltsanalyse findet sich in Abschnitt 2.2.5, oben.

Auch erste Versuche, auf der Grundlage der offenen Kommentare zur Pulsmessung eine Inhaltsanalyse und eine Sentiment Analyse durchzuführen, haben noch nicht überzeugt. Insbesondere deshalb, weil die Ergebnisse bisher nicht reproduzierbar sind.

GPT-4 hat auf der Grundlage des Prompts:

«Please perform the following: 1) Conduct a sentiment analysis for the comments in the file "Vorbereitung_Sentiment-Analyse_03_en_2.docx"; 2) display the results in an easy to understand way; 3) explain how you proceeded.»

beim ersten Versuch die nachfolgend gezeigten Ergebnisgrafiken erzeugt. Das sieht beeindruckend aus und passt auch in etwa zu den subjektiven Einschätzungen, die sich im Verlauf der Arbeit mit den Kommentaren ergeben haben.

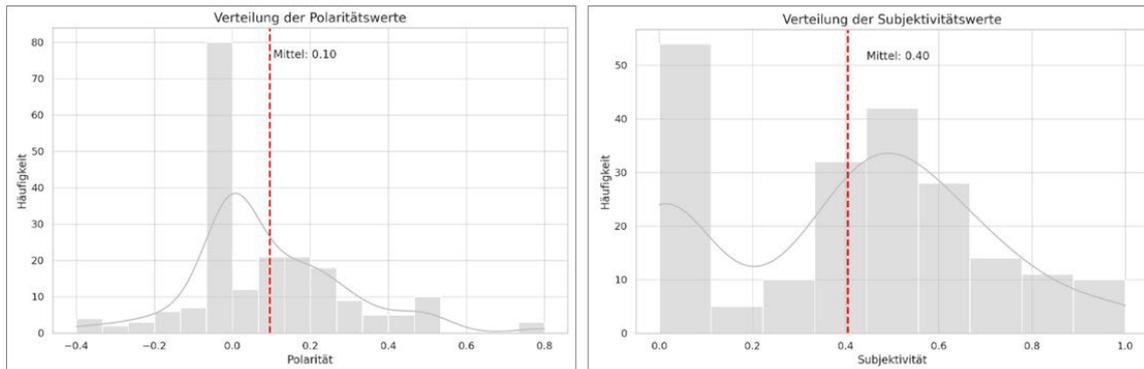


Abbildung 16: Sentiment-Analyse der offenen Kommentare zur Pulsmessung (Bildquelle: SCIL)

Leider gelang es auch mit wiederholten Versuchen nicht, diesen Detaillierungsgrad der Analyse und diese Qualität der Ergebnisdarstellung zu reproduzieren.

Das vorläufige Fazit aus diesen ersten Versuchen, mit GPT-4 qualitative Inhaltsanalysen und Sentiment Analysen zu den Kommentaren einer kurzen Befragung durchzuführen, ist daher folgendes: hier liegt noch ein offenes Feld für weiteres (Prompt-)Engineering vor.

3.2.7 Überprüfung der Ergebnisse

Ein wichtiger Aspekt der Arbeit mit Assistenzsystemen auf Basis generativer KI wie GPT-4 ist eine kritische Prüfung der Ergebnisse. Dabei sind verschiedene Verfahrensweisen möglich:

- stichprobenartige, manuelle Überprüfung von Zwischenschritten und / oder Endergebnissen;
- mehrfaches Anstossen der gleichen Erzeugung / Auswertung in separaten Dialogen bzw. Sessions;
- mehrfaches Anstossen der gleichen Erzeugung / Auswertung über identische oder leicht variierte Prompts;
- 'Chain-of-thought' prompting und Überprüfung von Zwischenschritten;
- Überprüfung des für die Datenanalyse erzeugten Python-Codes - entweder durch direkte Inspektion oder dadurch, dass dieser wiederum GPT-4 zur Analyse und Erläuterung vorgelegt wird;
- zusätzliche Auswertungen bzw. Nachberechnungen mit anderen Werkzeugen für Datenanalysen (z.B. SPSS, R, PowerBI, etc.)

Im Rahmen der Analysen zur Pulsmessung wurden folgende Prüfstrategien verwendet:

- (1) mehrfaches Anstossen von gleichen Auswertungen in separaten Dialogen bzw. Sessions und der Vergleich der Ergebnisse;
- (2) manuelle Nachberechnung eines einzelnen Werts auf jedem der ausgegebenen Ergebnisgraphen durch Filterung der jeweiligen Daten in MS Excel;
- (3) vollständige Überprüfung der Ergebniswerte für zwei ausgewählte Ergebnisgraphen (Abbildung 2 und Abbildung 3) auf Grundlage einer Auswertung der Daten mit SPSS.

Dabei hat sich gezeigt, dass bei (1) die erzeugten Ergebnisvisualisierungen und die darin eingetragenen Zahlenwerte ab einem bestimmten Reifegrad des Prompts reproduziert werden konnten; dass bei (2) die manuell errechneten Werte in allen Fällen mit den Werten der

Ergebnisgraphen von GPT-4 übereinstimmen; und dass bei (3) alle von SPSS ausgegebenen Ergebniswerte mit den von GPT-4 erzeugten Ergebniswerten übereinstimmen.

Allerdings gab es auch Überraschungen. Beim mehrmaligen Durchführen einer Auswertung zeigten sich zum Teil geringfügig unterschiedliche Zahlenwerte. Beispielsweise die Zahlenwerte für die Organisationen B und D in Abbildung 17 (152 und 288) und in Abbildung 18 (153 und 289). Oder die Zahlenwerte für die Altersgruppe 40-49 in Abbildung 17 (265) und in Abbildung 18 (266). Eine genauere Betrachtung zeigt, dass in Abbildung 17 Prozentwerte führend gezeigt werden, in Abbildung 18 absolute Häufigkeiten.

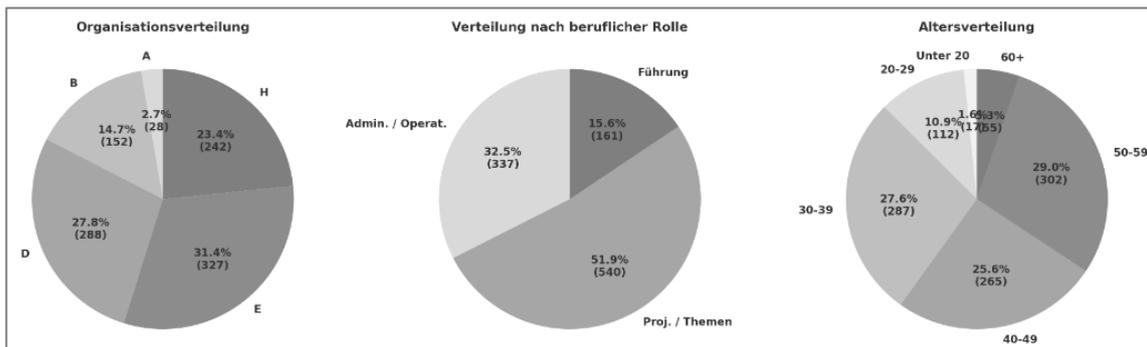


Abbildung 17: Verteilung der Teilnehmenden nach Organisation, Berufsrolle und Alter (1) (Bildquelle: SCIL)

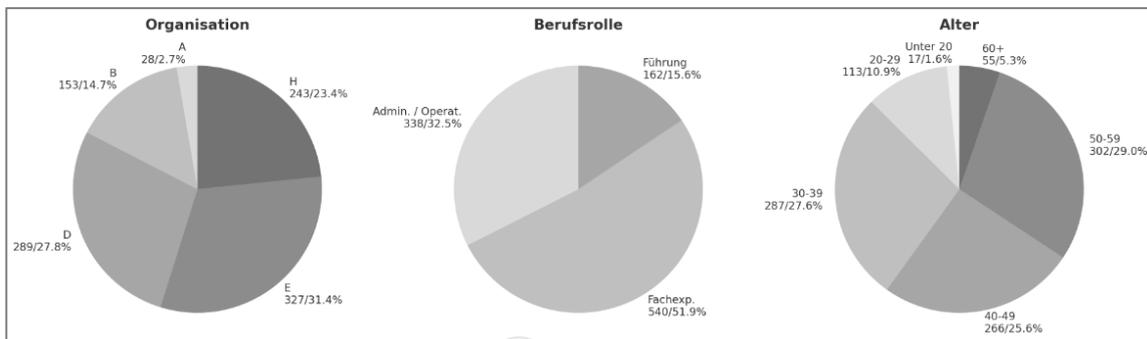


Abbildung 18: Verteilung der Teilnehmenden nach Organisation, Berufsrolle und Alter (2) (Bildquelle: SCIL)

Eine Vermutung war, dass für die Ausgabe in Abbildung 17 zunächst Prozentwerte berechnet wurden und davon ausgehend die Zahlenwerte für die absolute Häufigkeiten. Auf die Diskrepanz zwischen den Ergebnissen für die verschiedenen Durchläufe angesprochen, lieferte GPT-4 folgende Erläuterung:

The variation in figures, such as the number of respondents in a specific age group, could be due to several factors:

Rounding Differences: When calculating percentages and converting them back to absolute numbers, rounding can lead to minor discrepancies. For example, if a group constitutes 31.55% of 850 respondents, rounding the percentage to one decimal place results in 31.6%. The absolute number would then be either 268 or 269, depending on whether you round the product up or down.

Data Filtering or Processing: If there's any data filtering or processing applied inconsistently between analyses, it could result in slight differences. This might happen if there are missing or ambiguous entries in the data that are handled differently each time.

(...)

In der Folge wurde der Prompt zu dieser Auswertung so angepasst, dass die Bestimmung der absoluten Häufigkeit von Antworten im Vordergrund steht und erst danach die jeweiligen Prozentwerte berechnet werden:

«2) Please include (a) figures for exact numbers and (b) percentages in all segments shown.»

Diese Anpassung führte dann zum erwarteten und über die manuelle Prüfung bestätigten Ergebnis (vgl. Abbildung 1, oben).

3.3 ChatGPT&Co als Assistenten für Datenanalysen

Bildungsverantwortliche sind für Aufgaben im Bereich Datenanalysen und Learning Analytics nicht immer gut ausgebildet (vgl. Garavan et al. 2020). Hierbei sind zwei Aspekte zu unterscheiden: zum einen die Fähigkeit, zielführenden Fragen an verfügbaren Daten zu stellen; zum anderen die Fähigkeit, auf der Grundlage verfügbarer Daten Analysen durchzuführen und Ergebnisvisualisierungen zu erstellen. Diese Ergebnisvisualisierungen dienen dann wiederum als Grundlage für nachfolgende Diskussionen und Entscheidungsfindungen.

Die Pulsmessung, über deren Durchführung und Auswertung hier berichtet wurde, kann als Inspiration für datenbasiertes Entscheiden von Bildungsverantwortlichen in anderen Unternehmen und Organisationen dienen. Und zwar in zweierlei Hinsicht:

- (1) Die Pulsmessung ermöglichte eine schnelle Standortbestimmung zu Aspekten wie:
 - «Welche GenKI-Werkzeuge sind bei uns bekannt und werden genutzt?»,
 - «Wie wird die Zusammenarbeit mit diesen GenKI-Werkzeugen erlebt?»,
 - «Welche Unterstützung wünschen unsere Beschäftigten?».
- (2) Die Unterstützung der Auswertung einer Pulsmessung durch Werkzeuge wie GPT-4 eröffnet neue Möglichkeiten im Hinblick auf die Umsetzung von Datenanalysen und Ergebnisvisualisierungen.

Welche Wege für Bildungsverantwortliche sinnvoll sind, um Auswertungen bzw. Ergebnisvisualisierungen zu (kurzen) Erhebungen umzusetzen, hängt vom jeweiligen Kontext ab: von der Vertrautheit mit Konzepten und Methoden der empirischen Sozialforschung, von Erfahrungen mit den erforderlichen Arbeitstechniken, von der Verfügbarkeit geeigneter Werkzeuge (z.B. MS Excel, SPSS, MS-PowerBI, etc.) und von der Fingerfertigkeit im Umgang mit diesen Werkzeugen. Mit der Verfügbarkeit von smarten Assistenzsystemen wie ChatGPT&Co ergeben sich hier neue Optionen.

4 Implikationen für Verantwortliche in Bildung und Personalentwicklung

Die Ergebnisse der Pulsmessung und der Auswertung der Ergebnisse haben folgendes gezeigt:

- (1) das Wissen von Beschäftigten zu Werkzeugen bzw. Assistenzsystemen auf Basis generativer KI ist lückenhaft;
- (2) die Wahrnehmung der Nutzung von Werkzeugen generativer KI unterscheidet sich nach Alter und nach Beschäftigtengruppen.
- (3) für Bildungsverantwortliche eröffnen sich neue Möglichkeiten, datenbasiert zu arbeiten bzw. den Aufwand für Datenanalysen und Ergebnisvisualisierungen zu reduzieren.

Was folgt daraus?

- Verantwortliche für Bildung und Personalentwicklung sollten dazu beitragen, dass organisationsweite Initiativen zur Information zu Assistenzsystemen auf Basis generativer KI, zu Einsatzmöglichkeiten, zu Potenzialen und zu Limitationen auf den Weg gebracht werden.
- Verantwortliche für Bildung und Personalentwicklung sollten im Blick behalten, dass die Wahrnehmungen zu Assistenzsystemen auf Basis generativer KI und zum kollaborativen Arbeiten mit diesen Systemen unterschiedlich ausfallen können. Unterschiedliche Altersgruppen und unterschiedliche Beschäftigtengruppen haben unterschiedliche Sichtweisen auf die damit verbundenen Möglichkeiten und Herausforderungen. Dies muss bei der Konzeption und Umsetzung von Programmen zur Information und Befähigung von Beschäftigten berücksichtigt werden. Es ist sinnvoll, der Konzeption und Umsetzung von solchen Programmen Pulsmessungen vorzuschalten, um die Zielgruppen passgenauer bedienen zu können.
- Verantwortliche für Bildung und Personalentwicklung sollten die Möglichkeiten, die Assistenzsysteme auf Basis generativer KI im Hinblick auf Datenanalysen und Ergebnisvisualisierungen bieten, aktiv erkunden und gegebenenfalls für sich selbst nutzen. Auch für sie selbst ergeben sich aus der Verbreitung von ChatGPT&Co am Arbeitsplatz Entwicklungserfordernisse und Entwicklungschancen. Nicht zuletzt im Hinblick auf ein Bildungs- bzw. PE-Management, das weniger auf Bauchgefühl und anekdotische Evidenz setzt, sondern auf Datenbasierung.

Literaturverzeichnis

- Cheng, L., Li, X., & Bing, L. (2023). Is GPT-4 a good data analyst? ArXiv. Retrieved from <https://arxiv.org/pdf/2305.15038.pdf>
- Cook, V., & Bowley, T. Me, myself and AI: What you need to know. Bank of America Institute. <https://workplaceinsights.bofa.com/articles/2023/05/ai.html>
- Deepakvraghavan (2023). The hottest new programming language is English...And a picture is worth a thousand words! Beautiful GitOps architectures come to life. <https://deepakvraghavan.medium.com/the-hottest-new-programming-language-is-english-and-a-picture-is-worth-a-thousand-words-5e7a51ffb9>
- Einola, K., & Khoreva, V. (2023). Best friend or broken tool? Exploring the co-existence of humans and artificial intelligence in the workplace ecosystem. *Human Resource Management*, 62(1), 117–135. <https://doi.org/10.1002/hrm.22147>
- Garavan, T. N., Heneghan, S., O'Brien, F., Gubbins, C., Lai, Y., Carbery, R., . . . Grant, K. (2020). L&D professionals in organizations: Much ambition, unfilled promise. *European Journal of Training and Development*, 44(1), 1–86. <https://www.emerald.com/insight/content/doi/10.1108/EJTD-09-2019-0166/full/html>
- Gmyrek, P., Berg, J., & Bescond, D. (2023). Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality. ILO Working Paper 96. Geneva. <http://dx.doi.org/10.2139/ssrn.458421>
- Gruber, P. H. (2023). Using ChatGPT for Data Analysis without Programming. Instats Online Kurs, Oktober 2023.
- Kreacic, A., Uribe, L., Romeo, J., Lasater-Wille, A., Jesuthasan, R., & Luong, S. (2024). How generative AI is transforming business and society: The good, the bad, and everything in between. *oliverwymanforum.com*. <https://www.oliverwymanforum.com/global-consumer-sentiment/how-will-ai-affect-global-economics/workforce.html>
- Krugmann, J. O., & Hartmann, J. (2024). Sentiment analysis in the age of generative AI. *Customer Needs and Solutions*, 11(3). <https://doi.org/10.1007/s40547-024-00143-4>
- Morgan, D. L. (2023). Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT. *International Journal of Qualitative Methods*, 22, 16094069231211248. <https://doi.org/10.1177/16094069231211248>
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192. <https://doi.org/10.1126/science.adh2586>

Seufert, S., & Handschuh, S. (2024). Generative KI: Mensch-Maschine-Augmentation. In S. Seufert & S. Handschuh (Hrsg.), *Generative Künstliche Intelligenz: ChatGPT und Co für Bildung, Wirtschaft und Gesellschaft*. Stuttgart: Schäffer-Poeschel.

SWK (2024). Large Language Models und ihre Potenziale im Bildungssystem. Impulspapier der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz. [www.swk-bildung.org. https://www.kmk.org/fileadmin/Dateien/pdf/KMK/SWK/2024/SWK-2024-Impulspapier_LargeLanguageModels.pdf](https://www.kmk.org/fileadmin/Dateien/pdf/KMK/SWK/2024/SWK-2024-Impulspapier_LargeLanguageModels.pdf)

Yosifova, A. (2024). ChatGPT Code Interpreter: What It Is and How It Works. Retrieved from <https://365datascience.com/trending/chatgpt-code-interpreter-what-it-is-and-how-it-works/#1>

Zhang, H., Wu, C., Xie, J., Lyu, Y., & Cai, Jie, Carroll, John M. (2023). Redefining qualitative analysis in the AI era: Utilizing ChatGPT for efficient thematic analysis. *ArXiv*. <https://arxiv.org/abs/2309.10771>

Anhang

Im Rahmen der Auswertung dieser Pulsmessung zum Thema «Arbeiten mit generativer KI» wurden die von den Befragten beigesteuerten Kommentare auch ohne technische Hilfe durch GPT-4 durchgesehen. Im Folgenden einige ausgewählte Aussagen, die im Kommentarfeld zur Frage nach dem Erleben der Arbeit mit GenKI-Werkzeugen eingegeben wurden. Diese sind nicht kategorisiert oder sortiert, sondern werden in der Reihenfolge ihres Erscheinens wiedergegeben. Diese Zusammenstellung soll einen Eindruck von der Vielfalt der Sichtweisen vermitteln.

Ausgewählte Kommentare zum Erleben der Zusammenarbeit mit generativer KI

#22, Q19

«Ich frage mich nicht, ob es mich und meine Arbeitskraft noch braucht, sondern ob es mich und meine Arbeitskraft in dieser Form noch braucht.»

#51, Q19

«Ich denke, dass das Vorausdenken sowie Mitdenken dadurch verloren geht. Man muss mit der KI-Hilfsmitteln meistens nicht viel überlegen und wissen, sondern lediglich den Tool nutzen können. Das macht mir Sorgen für die Zukunft. Ich finde, es ist ein tolles Hilfsmittel zum Arbeiten, aber ab einen gewissen Punkt nimmt uns unsere "geistige Entwicklung" weg.»

#79, Q19

«Es ist ein mächtiges Werkzeug, für mich besteht die größte Gefahr darin, dass es die Nutzer geben wird, die Abkürzungen nehmen, um sich einen Vorteil gegenüber anderen "normalen" Nutzern zu verschaffen, die es ordnungsgemäß verwenden. Diejenigen, die es nicht nutzen, werden im Vergleich zu den Nutzern an Effektivität verlieren.» (Übersetzt aus dem Französischen mit DeepL.com)

#96, Q19

«Ich fühle mich vor allem gestärkt, da ich verschiedene Aufgaben effizienter bewältigen kann.»

#98, Q19

«Es hat wie üblich zwei Seiten, zum einen ergibt es sicherlich eine Erleichterung die letzten entscheidenden Prozent muss man aber selber beisteuern bzw. Quellen prüfen von Texten, ob das wirklich so stimmig ist usw. aber sicher bei der Codegenerierung bin ich immer wieder erstaunt und das nimmt bereits viel Kopfarbeit ab wird aber dann in das Verständnis transferiert»

#173, Q19

«Die KI ist jeweils nur so gut, wie der Nutzer des Werkzeuges. Somit ist es wichtig zu Wissen, wie diese Werkzeuge effektiv und effizient genutzt werden können. Dies schliesst aus, dass es sich bei der Nutzung von KI wie schummeln anfühlt. Hierbei handelt es sich eher darum die Fähigkeiten präzise einzusetzen.»

#470, Q19

«Also bisher hätte ich kein Aufgabengebiet gefunden, wo genAI etwas unterstützt, was ich sonst nicht geschafft hätte. Viele Sachen, die das täten wie z.B. M365 Co-Piloting nutzen wir ja leider noch gar nicht. Dadurch würde ich mir wirklich eine Produktivitätssteigerung erwarten. Hoffe, es kommt im Q4»

#575, Q19

«Ich habe chatgpt im Rahmen einer Bachelor-Ausbildung, die ich neben meiner Arbeit mache, ausprobiert. Das Tool ist interessant, aber meiner Meinung nach fehlt es an Präzision bei den Antworten. Zumindest in meinem Ausbildungs- und Tätigkeitsbereich (Recht).» (Übersetzt aus dem Französischen mit DeepL.com)

#638, Q19

«Especially CoPilot sometimes provides wrong / useless suggestions. then it costs me more time because it interrupts the flow of coding. when I don't know about a solution I still have to look it up online to see what the problem was and why it works, if I want robust code and not just "works at the moment, let's leave it".
the time gain is only if it adds closing brackets and stuff like that that is annoying to type.»

#880, Q19

«Wo führt das Ganze hin?

Wie entwickelt sich das weiter?

Wie geht die Gesellschaft damit um? Gibt es eine Zweiteilung, solche die es kennen und damit umgehen können und solche die damit gar nichts anfangen können?»

#902, Q19

«Die KI hat nicht die Sensibilität eines Menschen und reproduziert nur, was ihr gezeigt wird. Sie erschafft nichts, sie wird nicht in der Lage sein, ein Problem mit innovativen Lösungen oder außerhalb der bereits etablierten Normen zu bewältigen.» (Übersetzt aus dem Französischen mit DeepL.com)

#941, Q19

«Es gibt vermehrt wirklich nützliche KI basierte Tools, aber es wird auch viel Mist produziert. Es wird meist zu viel versprochen und die Realität sieht man dann wieder eher nüchterner. Die meisten Menschen können ohne Ausbildung zum Thema gar nicht begreifen, um was es geht. Sie denken eher an Science Fiction Szenarien (gut sowie auch böse) als an nützliche Werkzeuge. (...) Ich finde nicht gut, dass der Fokus (und die Budgets) auf generative KI gelegt wird, und viele andere potentielle KI gestützte Tools (wie z.B. Security Data Analytics), welche meiner Meinung nach nützlicher sind, wenig beachtet werden und nicht die Budgets bekommen, welche dem jeweiligen Thema angemessen wären.»

#1040, Q19

«GenAI wird uns (hoffentlich) noch viele interessante, effizienzsteigernde Lösungschancen bieten. In der aktuellen "Wild-West-Phase" sollte man meiner Meinung nach so viel wie möglich

damit ausprobieren - mit dem nötigen gesunden Menschenverstand (bzgl. privaten/vertraulichen Daten). Ich bin überzeugt, dass es uns weiterhin benötigt und GenAI uns nicht obsolet machen wird. Viel mehr freue ich mich auf all die Ressourcen, welche uns GenAI "freischaufeln" wird, die wir wiederum für die spannenden, vielfältigen Tasks einsetzen können.»