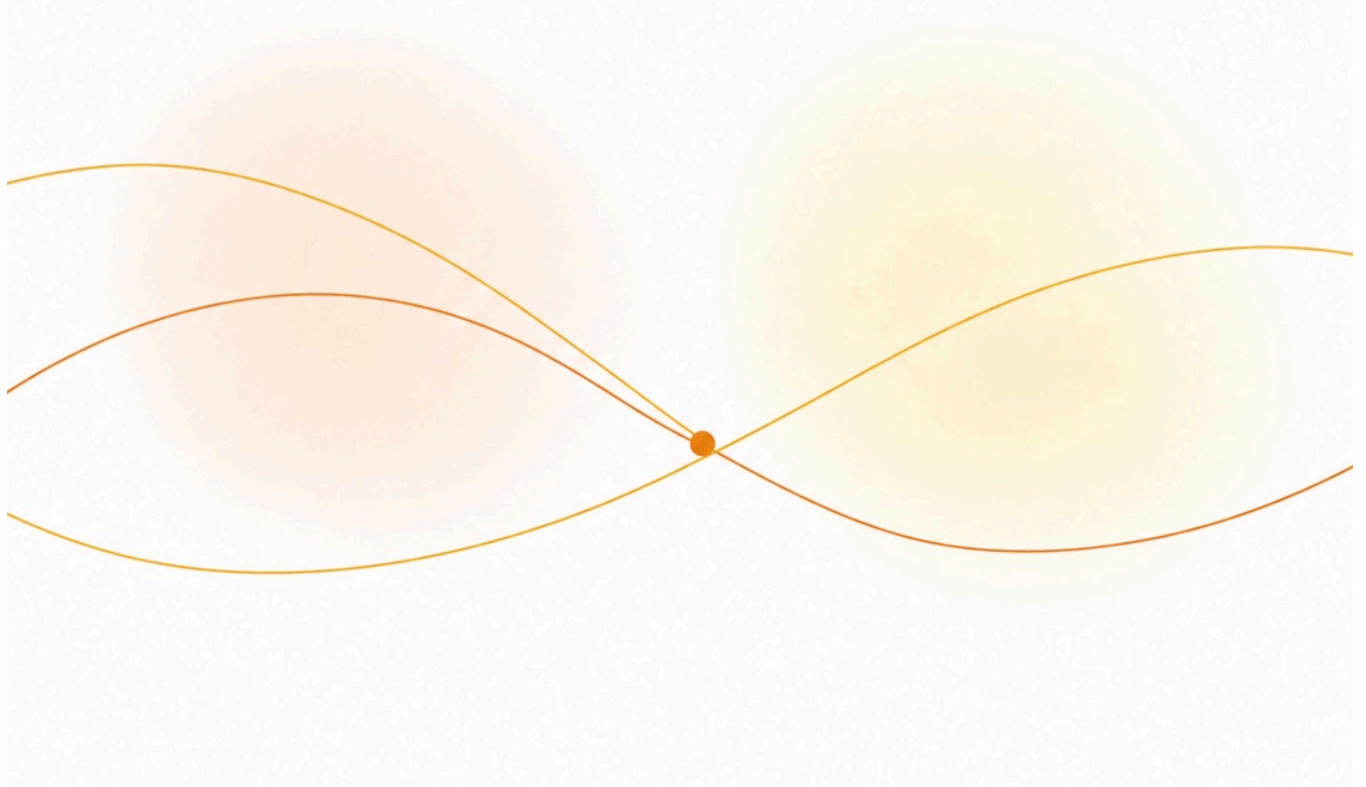


WORKSHOP WEAVER · STUDIE

State of AI Facilitation 2026

Was KI-Assistenten
Facilitator:innen empfehlen, und
was in ihren Workshop-Agenden
fehlt

Neun KI-Agenden und 720 Antworten von KI-Assistenten: Die Zeiten
stimmen, die Warnhinweise zu den Methoden fehlen. Worauf du vor
dem Workshop achten musst.



1. Was haben wir gefunden?

Im Frühjahr 2026 veränderte sich eine Frage, die uns Facilitator:innen immer wieder stellen. Bis dahin wollten sie wissen, ob sie einen Workshop überhaupt mit KI planen sollten. Seitdem wollen sie wissen, wofür sie ihr trauen können. Wir entwickeln Workshop Weaver, ein Werkzeug für die Workshop-Planung, und hatten dazu eine Meinung, aber keine Daten. Dieser Report liefert sie nach. Er wertet 720 Antworten aus, die drei KI-Assistenten auf Fragen gaben, wie Facilitator:innen sie tatsächlich stellen, und bewertet 9 Agenden, die dieselben Assistenten entworfen haben. Das Bewertungsraster orientiert sich daran, was eine erfahrene Facilitatorin vor einem Workshop prüft. Die Stichprobe ist klein, und wir verstehen die Ergebnisse als erstes Signal. Zu jeder Zahl steht die Grundgesamtheit, und jede lässt sich in den veröffentlichten Daten nachvollziehen.

- **Bei acht von neun Agenden in unserem Test ergaben die Blöcke mit Pausen genau die Workshoplänge. Dieselben acht hielten alle festen Vorgaben des Briefings ein.**

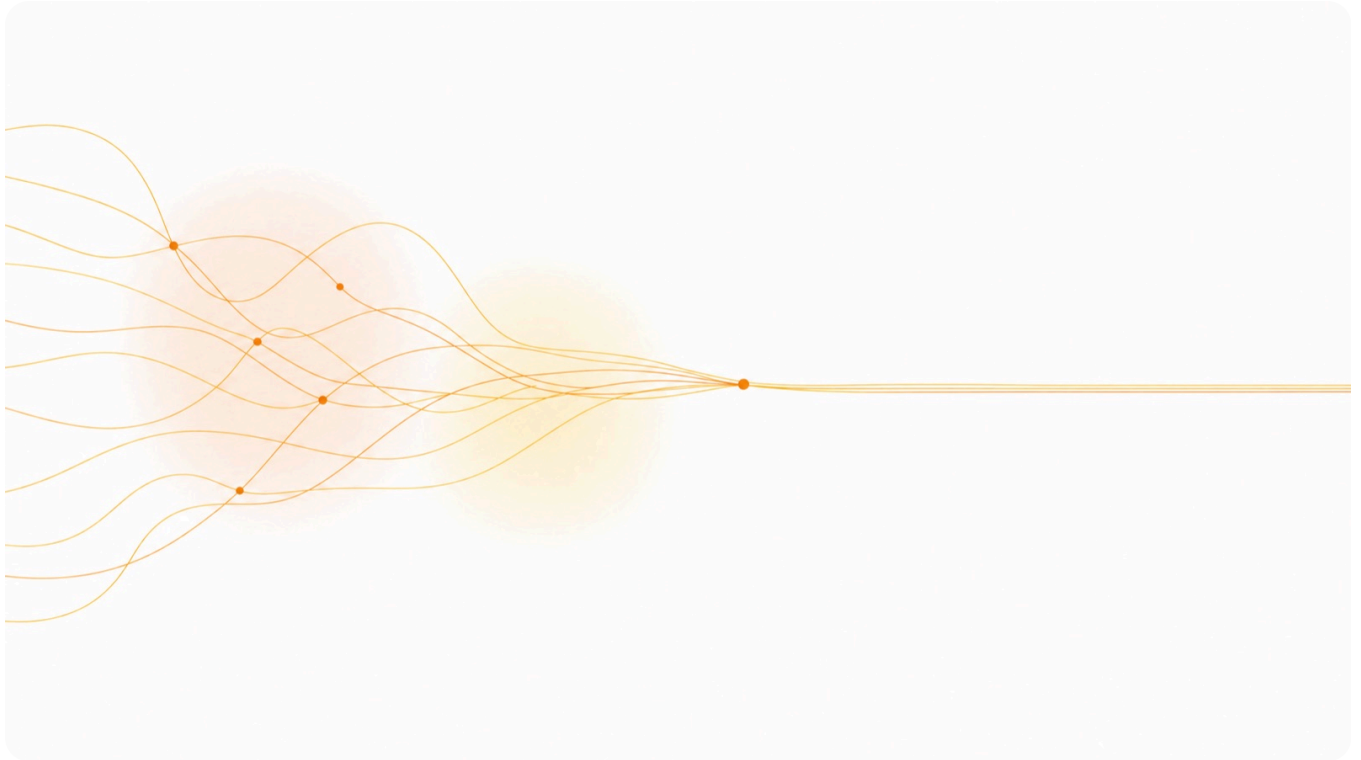
- **Keine der neun Agenden hielt fest, woher eine Methode stammt, obwohl alle neun mit gängigen Methoden arbeiteten.**

- **Sieben von neun Agenden setzten eine Methode in einer Situation ein, vor der ihre Beschreibung warnt, ohne darauf hinzuweisen. In fünf Fällen war es ein Dot Voting, das im Beisein der Führung eine verbindliche Entscheidung vorbereitete.**

- **Auf 108 Anfragen nach einer Agenda kam ein einziges Mal eine Rückfrage zurück. Bei der absichtlich vagen Anfrage war es eine von 18.**

- **Sechs Methoden machen die Hälfte aller 541 Methodennennungen aus. Dot Voting kommt in 114 von 270 Antworten zu Methoden und Agenden vor.**

In den Kapiteln 2 und 3 stehen die Daten hinter diesen Befunden. Kapitel 4 zeigt an einem einzelnen Agenda-Block, wie die Lücke auf dem Papier aussieht. Kapitel 5 enthält die fünf Fragen, die wir inzwischen an jede KI-Agenda stellen. In Kapitel 6 beschreiben wir unser Vorgehen und seine Grenzen, angefangen bei der Tatsache, dass eine KI die Agenden bewertet hat.



2. Was empfehlen die Assistenten, wenn eine Facilitatorin fragt?

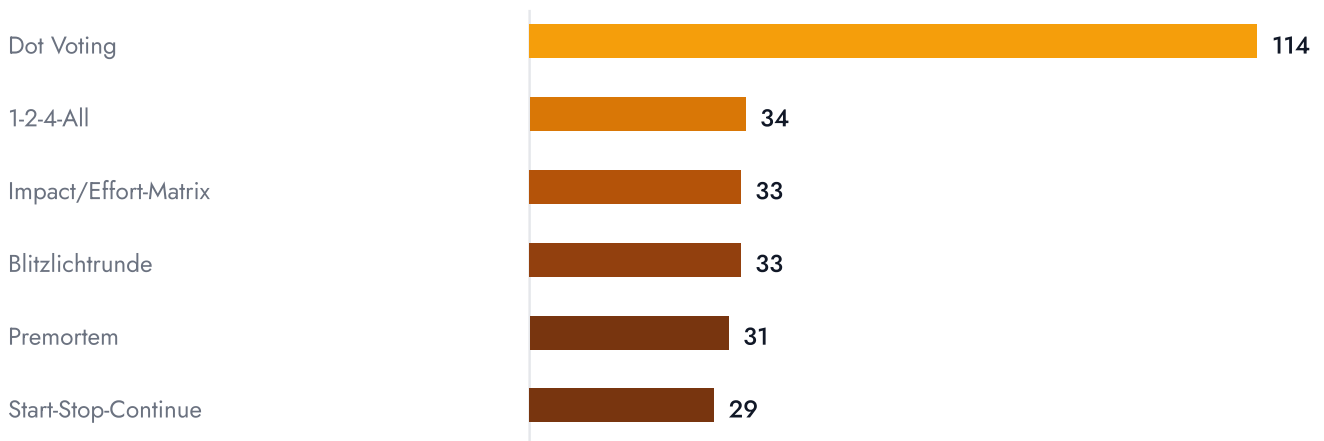
Sie empfehlen immer wieder dieselben sechs Methoden, und sie liefern lieber einen Plan als eine Rückfrage. Für diese Untersuchung haben wir Claude, ChatGPT und Perplexity 40 Fragen gestellt, jede auf Deutsch und auf Englisch, jede dreimal und stets mit eingeschalteter Websuche. So kamen 720 Antworten zusammen. Wir haben gezählt, welche Werkzeuge und Methoden darin genannt werden, und jede zitierte Website danach eingeordnet, wer sie betreibt. Die Regeln dafür veröffentlichen wir zusammen mit den Rohdaten.

Methoden: immer dieselben sechs Namen

Sechs Methoden machen die Hälfte aller 541 Methodennennungen aus: Dot Voting, 1-2-4-All, die Impact/Effort-Matrix, Round Robin, das Premortem und Start-Stop-Continue. Dot Voting allein erscheint in 114 von 270 Antworten zu Methoden und Agenden. Das ist ein brauchbarer Methodenkoffer, mit dem sich die meisten Workshops bestreiten lassen. Eine Prüfung fiel sogar besser aus, als wir erwartet hatten. Nannte eine Frage die Gruppengröße, enthielten nur 4 von 144 Antworten eine Methode, die laut unserer Methodensammlung nicht zu dieser Gruppengröße passt.

Deutlich seltener erfährt man aus den Antworten, wann eine Methode nicht funktioniert. Auf die Frage, wann man Dot Voting besser nicht einsetzt, gingen nur 5 von 18 Antworten überhaupt auf die Gruppengröße ein, obwohl unsere Beschreibung sie als erste Einschränkung nennt. In 28 von 72 Antworten auf unsere Strategiefragen empfahlen die Assistenten Dot Voting oder ein Premortem für einen Workshop, in dem verbindliche Entscheidungen anstanden. Genau davor warnen beide Methodenbeschreibungen.

Abbildung 1 · Sechs Methoden machen die Hälfte aller Methodennennungen aus



Quelle: unsere Befragung von KI-Assistenten, September 2026. 270 Antworten auf unsere Methoden- und Agendafragen.
Gezählt: Antworten, die die Methode nennen.

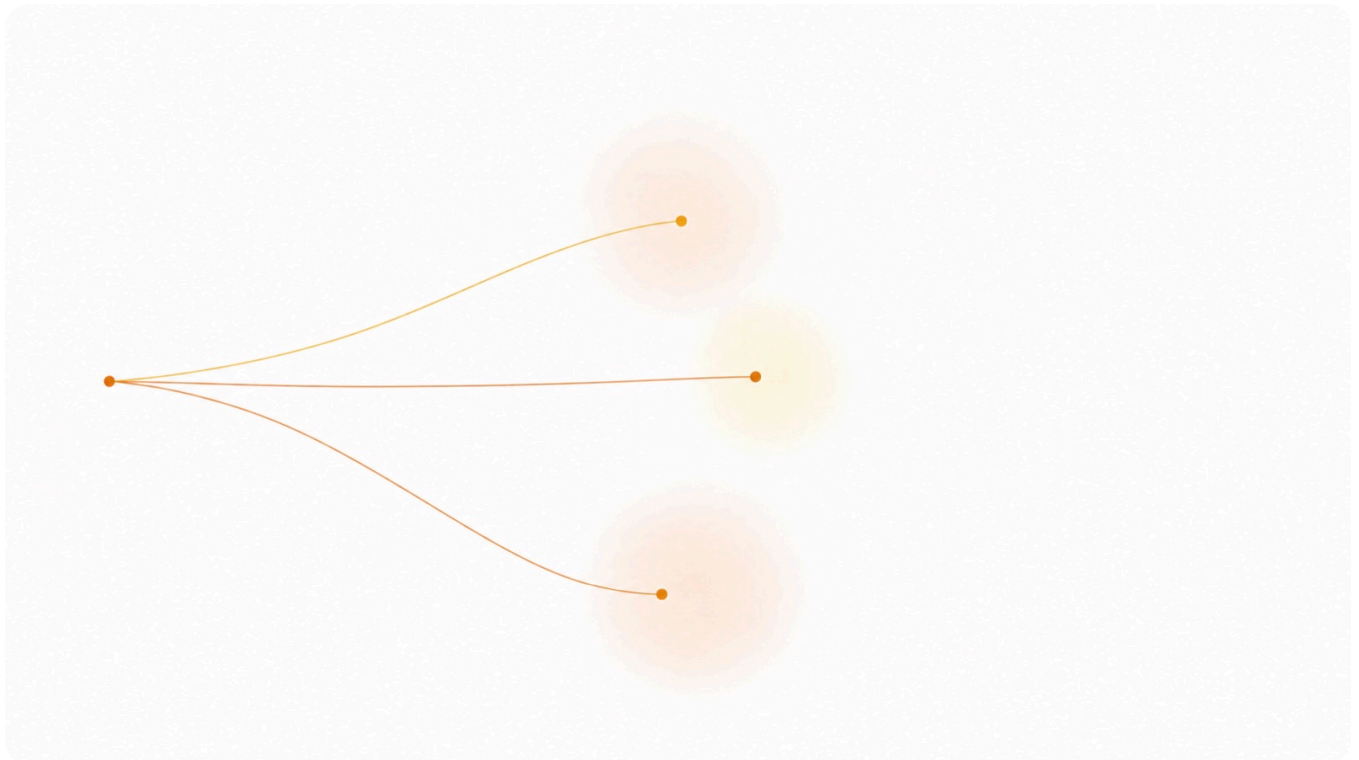
Agenden: eine Rückfrage in 108 Antworten

Wer einen KI-Assistenten um eine Agenda bittet, bekommt eine. Von 108 Antworten auf Agenda-Anfragen hielt 1 den Plan zurück und stellte zuerst eine Frage. Selbst auf die absichtlich vage Anfrage („Ich brauche eine halbtägige Retrospektive für mein Team. Das Quartal lief schlecht.“) kam nur in 1 von 18 Antworten eine Rückfrage. Alle anderen lieferten eine vollständige Agenda für ein Team, dessen Größe und Arbeitsweise sie selbst angenommen hatten. Viele legten ihre Annahmen offen, und das ist die ehrlichere Variante. Eine Rückfrage wäre trotzdem etwas anderes gewesen.

Werkzeuge und Quellen: die Kurzfassung

Auf die Frage nach einem Werkzeug für die Workshop-Planung nannten die Assistenten in 127 von 180 Antworten ein Produkt, das eigens dafür entwickelt wurde. An erster Stelle stand in 83 von 180 Antworten SessionLab; Workshop Weaver tauchte in 27 von 180 Antworten auf. Von den 6.349 zitierten Websites, die wir zuordnen konnten, gehören 49,7 Prozent einem Unternehmen, das ein Werkzeug oder eine Dienstleistung verkauft, und nur 14,7 Prozent einer Facilitatorin oder einem Facilitator.

In den Antworten auf die deutschen Fragen kamen zudem kaum Anbieter aus dem eigenen Markt vor: 62 von 90 deutschen Antworten auf die Werkzeugfragen nannten kein einziges Produkt aus dem deutschsprachigen Raum. Stattdessen stand in 27 von 90 deutschen Antworten ein allgemeiner KI-Chatassistent an erster Stelle, in den englischen nur in 8 von 90. Daraus ergibt sich eine einfache Regel: Schau nach, wem die Seite gehört, die der Assistent als Quelle nennt.



3. Wie gut planen die Assistenten einen Workshop?

Die Zeitpläne stimmten, und die Agenden hielten die Vorgaben aus dem Briefing ein. Was fehlte, waren die Warnhinweise zu den Methoden. Für den Test haben wir 3 Briefings geschrieben: eines auf Englisch mit allen nötigen Angaben, eines auf Deutsch mit fester Mittagspause und einem Gastbeitrag der Geschäftsführung, und ein drittes absichtlich vage gehalten. Jedes Briefing ging einmal an Claude, ChatGPT und Perplexity, diesmal ohne Websuche. So entstanden 9 Agenden. Die Arbeitsanweisung verlangte einen vollständigen Ablauf mit Zeiten, einer Methode für jeden Block, den Aufgaben von Moderation und Teilnehmenden sowie dem Material. Nach Quellen, nach Warnhinweisen zu den Methoden oder nach Rückfragen verlangte sie nicht. Die Ergebnisse zeigen deshalb, was die Assistenten von sich aus liefern.

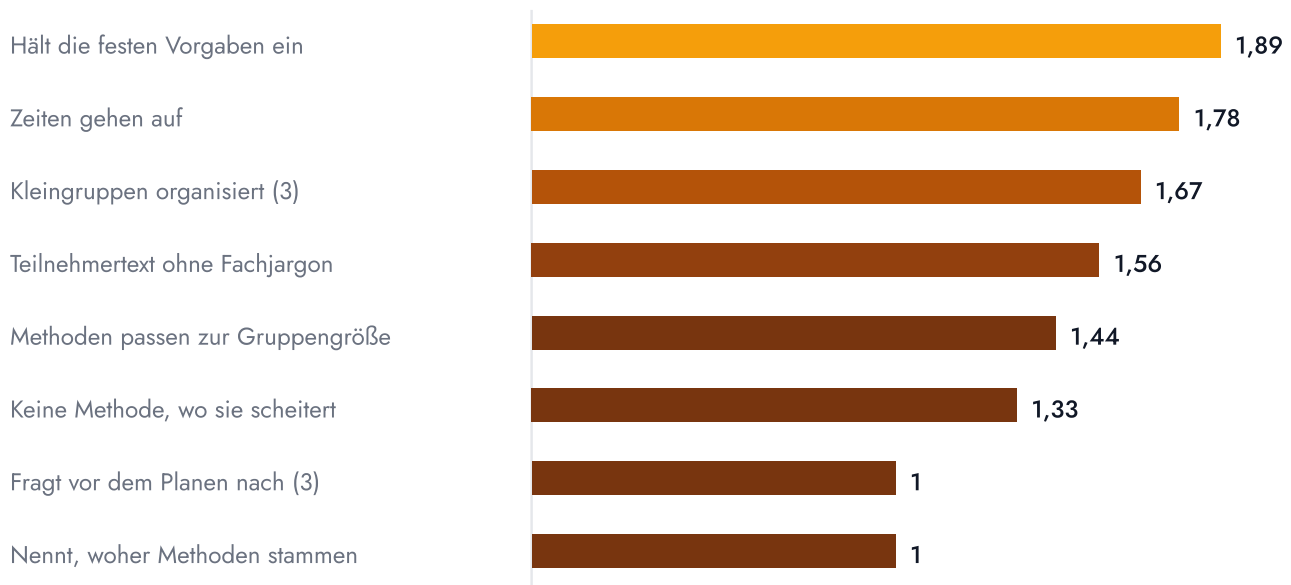
Bewertet hat Claude, also eine KI, die selbst zu den getesteten Assistenten gehört. Die Bewertung lief verdeckt und nach einem schriftlichen Raster mit acht Kriterien, für die es jeweils 0 (fehlt), 1 oder 2 Punkte (voll erfüllt) gab. Kein Mensch hat die Agenden bewertet. Eine zweite, unabhängige Bewertung stimmte mit der ersten in 65 Prozent der 40 Fälle überein, weniger als die 70 Prozent, die wir uns vorgenommen hatten. Die Punkte sind deshalb als maschinelle Lesart nach einem festen Raster zu verstehen und nicht als Urteil einer Facilitatorin. Kapitel 6 beschreibt das Verfahren im Einzelnen.

Was die Agenden richtig machten

Beim Zeitplan und beim Briefing gab es wenig zu beanstanden. Bei 8 von 9 Agenden ergaben die Blöcke mit Pausen genau die Workshoplänge, und jeder Block hatte eine Zeitangabe. Dieselben acht hielten alle festen Vorgaben ein, im deutschen Briefing also die Mittagspause um halb eins, den Gastbeitrag ohne anschließende

Diskussion, das pünktliche Ende und die Anrede mit Sie. Die neunte Agenda, Perplexitys Antwort auf das vage Briefing, kündigte einen Workshop von vier Stunden an und plante dann nur drei. In dem Briefing, das Kleingruppen verlangte, beschrieben 2 von 3 Agenden die Gruppen, die Aufteilung, das Timing und die Rückmeldung ins Plenum. Von den 60 Wertungen für alle neun Agenden erreichten 29 von 60 die volle Punktzahl, nur 1 von 60 lag bei null.

Abbildung 2 · Die Agenden halten Briefing und Zeiten ein; selten sagen sie, woher eine Methode stammt oder wann sie scheitert



Quelle: unser Agenda-Test, September 2026. Neun Agenden (drei Workshop-Briefings, jeweils an Claude, ChatGPT und Perplexity), je Kriterium mit 0 (fehlt) bis 2 (voll erfüllt) bewertet; gezeigt ist der Durchschnitt. Kleingruppen nur beim Briefing bewertet, das sie verlangte, Nachfragen nur beim vagen Briefing (je drei Agenden).

Was in den Agenden fehlte

Es fehlten die Herkunft der Methoden und die Hinweise darauf, wann sie nicht passen. Keine der neun Agenden nannte, woher eine Methode stammt. Die Methoden wurden benannt, aber weder zugeordnet noch belegt, was in unserem Raster bei jeder Agenda einen von zwei Punkten bedeutet. Eine Facilitatorin kann eine solche Methode nur anhand der Beschreibung prüfen, die der Assistent selbst mitliefert.

0

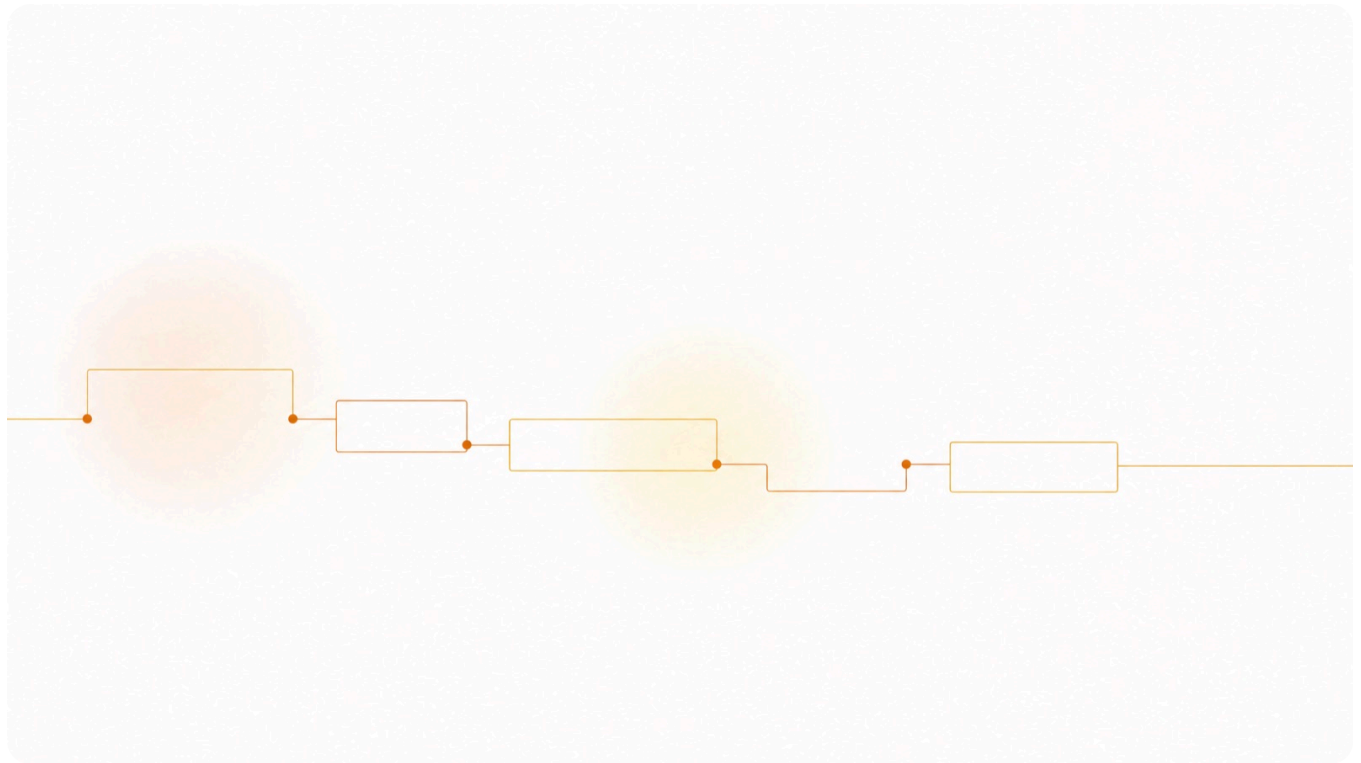
von 9 Test-Agenden
nennen die Herkunft
einer Methode

Die zweite Lücke waren die Warnhinweise. In 7 von 9 Agenden kam eine Methode in einer Situation zum Einsatz, vor der ihre Beschreibung in unserer Sammlung ausdrücklich warnt, und keine dieser Agenden erwähnte das. Fünfmal handelte es sich um ein Dot Voting, das im Beisein der Führung eine verbindliche Entscheidung vorbereitete oder herbeiführte. Zweimal wurde eine Clustering-Übung deutlich schneller angesetzt, als ihre Beschreibung erlaubt, einmal davon in Kombination mit einer Check-in-Runde, die nur ein einziges Wort zuließ. Die Gruppengröße war in 4 von 9 Agenden bei jedem Block angegeben. Der Text für die Teilnehmenden war in 5 von 9 Agenden sauber von den Moderationsnotizen getrennt.

Das vage Briefing

Auf das englische Briefing („I need a half-day retrospective for my team. We had a rough quarter.“) reagierte keiner der drei Assistenten mit einer Rückfrage. Jeder nannte stattdessen in einer Zeile seine Annahme, vier Stunden und ein Team von sechs bis zwölf Personen, und begann zu planen. Unsere Arbeitsanweisung hatte genau dieses Verhalten nahegelegt. Schwerer wiegt deshalb der Befund aus Kapitel 2: Dort gab es keine solche Anweisung, und trotzdem enthielt nur 1 von 18 Antworten eine Rückfrage. Keiner der drei Assistenten stellte die Frage, mit der eine Facilitatorin anfangen würde: Was ist eigentlich passiert?

Ob die Assistenten Quellen, Warnhinweise oder Rückfragen liefern, wenn man sie ausdrücklich darum bittet, hat dieser Test nicht untersucht. Er zeigt aber, an welcher Stelle sich das genaue Lesen einer KI-Agenda lohnt. Die Zeitrechnung stimmte in acht von neun Agenden; die Methodenwahl ist der Teil, den du selbst prüfen musst. Worauf es dabei ankommt, zeigt das nächste Kapitel an einem einzelnen Block.



4. Woran erkennt man auf dem Papier, dass ein Dot Voting schiefgehen wird?

An einer guten Agenda. Der folgende Block stammt aus Claudes Antwort auf das deutsche Strategie-Briefing: zehn Personen aus einem Führungsteam, ein Tag, eine gemeinsame Drei-Jahres-Ambition und die ersten drei Initiativen dafür. Die Herkunft der Methoden blieb sie schuldig wie alle neun Agenden; bei jedem anderen Kriterium bis auf eines erreichte sie die volle Punktzahl. Gerade weil die Planung so gut ist, lässt sich die Lücke hier am besten zeigen.

Der Block, wie der Assistent ihn geschrieben hat

Die Agenda beginnt mit einer Annahme; im Block um 15:15 Uhr werden dann die drei Initiativen ausgewählt.

Die Geschäftsführung gehört zu den zehn Teilnehmenden und bleibt ganztägig anwesend; ...

Die Geschäftsführung sitzt also den ganzen Tag mit am Tisch. Um 15:15 Uhr heißt es dann:

Zweistufig – (1) Wirkung/Machbarkeit-Matrix: Cluster werden gemeinsam auf einer Vier-Felder-Matrix platziert (15 Min.); (2) Punktabfrage: je 3 Klebpunkte pro Person auf die Kandidaten im Feld „hohe

Wirkung“ (10 Min.); (3) Kurz-Check der drei Erstplatzierten auf Widerstand (5 Min.).

Mit der Punktabfrage ist das Dot Voting aus den Kapiteln 2 und 3 gemeint. Am Ende der Agenda steht unter den Notfallregeln eine Zeile, die man zweimal lesen sollte:

Zieht sich die Priorisierung: Punktabfrage entscheidet, die Matrix-Diskussion wird abgebrochen.

Wird die Zeit knapp, entscheiden also die Klebepunkte, und die Diskussion an der Matrix entfällt.

Wovor die Methodenbeschreibung warnt

Unsere Beschreibung von Dot Voting nennt vier Situationen, in denen man die Methode nicht einsetzen sollte. Zwei davon treffen auf diesen Block zu. Erstens ist die Entscheidung verbindlich und weitreichend, während Klebepunkte nur Vorlieben messen und nichts über Machbarkeit oder Belege sagen. Zweitens ist das Gefälle im Raum groß, und abgestimmt wird öffentlich, also genau die Situation, für die die Beschreibung ein anonymes Verfahren empfiehlt. Der Abschnitt „Stolperfallen“ derselben Beschreibung liefert eine dritte Schwierigkeit. Wer nacheinander klebt, orientiert sich an den Punkten, die schon hängen, und der Block legt nicht fest, ob alle gleichzeitig kleben. Die Notfallregel macht aus diesem Risiko schließlich den Plan, denn unter Zeitdruck entscheidet dann allein das Punktebild.

Die faire Lesart

Wer diesen Block verteidigt, hat gute Argumente. Läuft der Block wie geplant, treffen die Klebepunkte lediglich eine Vorauswahl. Die Matrix davor klärt die Machbarkeit, und das ist genau die Alternative, die unsere Beschreibung selbst vorschlägt. Die Widerstandsabfrage danach ist der Ort, an dem tatsächlich entschieden wird. Der Eröffnungsblock kündigt zudem Konsent als Entscheidungsmodus des Tages an. Unsere erste Bewertung folgte dieser Lesart und gab dem Block die volle Punktzahl. Die zweite las ihn als Grenzfall ohne Warnhinweis und vergab einen Punkt, und wir haben uns der zweiten angeschlossen. Einig sind sich beide Lesarten in einem: Die Bedingung und die Notfallregel stehen auf dem Papier, der Warnhinweis nicht.

Die Korrektur

Eine Facilitatorin braucht zwei Sätze, um den Block zu reparieren: „Alle kleben gleichzeitig, und die Geschäftsführung klebt zuletzt oder gar nicht. Das Punktebild ist eine Vorauswahl, entschieden wird in der Widerstandsabfrage.“ Dazu kommt eine Änderung in den Notfallregeln. Wird die Zeit knapp, entfällt die Diskussion an der Matrix, und die Widerstandsabfrage bleibt. Die Methode selbst kann bleiben. Der Warnhinweis muss nur dorthin, wo die Moderation ihn um 15:14 Uhr liest.

5. Was prüfst du, bevor du eine KI-Agenda durchführst?

Fünf Dinge, die je eine Minute kosten. Die Zeitrechnung gehört nicht dazu. Wie nötig diese Prüfung ist, zeigt, wie verbreitet KI in der Workshop-Planung inzwischen ist. Laut dem State of Facilitation 2026 von SessionLab (714 Antworten) nutzen inzwischen 38,7 Prozent der Facilitator:innen häufig generative KI. Die Ausgabe 2025 (1.050 Antworten) hatte ermittelt, dass 25,9 Prozent der KI-Nutzenden damit bereits komplette Agenden entwerfen lassen, also etwa jede vierte Person.

Unsere Suche in der veröffentlichten Forschung ergab 139 Studien zu KI in moderierten Gruppen. Alle arbeiteten mit Modellen von 2025 oder früher und sagen deshalb wenig über die heutigen Assistenten aus, wohl aber etwas über Gruppen.

Bei Alsobay und Kollegen (2026) teilten Gruppen unter KI-Moderation mehr von ihrem Wissen als Gruppen unter menschlicher Moderation, trafen aber nicht häufiger die beste Entscheidung. Bei Parisi und Kollegen (2026) verschob die KI-Moderation die Budgetverteilung der Gruppe, indem sie Vorschläge von Außenseitern neu formulierte. Eine Studie, die eine KI-Agenda mit der einer Facilitatorin vergleicht, haben wir nicht gefunden, ebenso wenig eine, die aktuelle Modelle in moderierten Gruppen testet. Alle Studien sind mit ihren Auszügen in den veröffentlichten Daten enthalten.

Die fünf Fragen

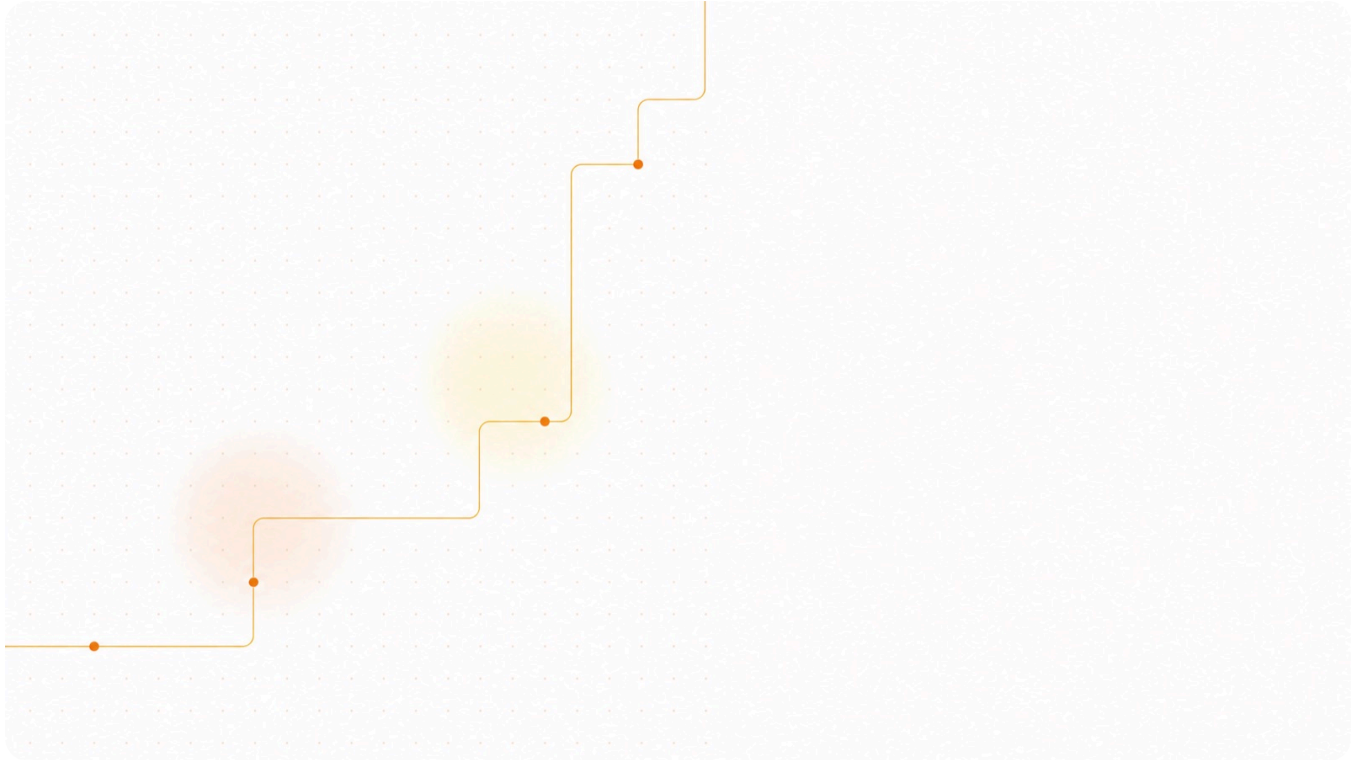
BEVOR DU EINE KI-AGENDA DURCHFÜHRST

1. Ergeben die Blöcke mit Pausen genau die Workshoplänge, und steht das so in der Agenda?
2. Für welche Gruppengröße ist jede Methode gedacht, und welche Gruppe hast du?
3. Gibt es zu jeder Methode ein „Wann nicht einsetzen“, und trifft es hier zu?
4. Versteht man den Text für die Teilnehmenden auch ohne Moderationsnotizen und Fachbegriffe?
5. Hatte das Briefing feste Vorgaben, und stehen sie genau dort, wo der Auftraggeber sie gesetzt hat?

Was das für 2027 bedeutet

Wir erwarten, dass allgemeine Assistenten bei Zeitplan und festen Vorgaben zuverlässig bleiben und dass Warnhinweise zu Methoden und Rückfragen weiterhin fehlen, solange die Anweisung sie nicht ausdrücklich verlangt. Beides sind Erwartungen, keine Befunde; sie stützen sich auf neun Agenden und einen einzigen Testtag. Die Ausgabe 2027 wird sie mit den dann aktuellen Modellen überprüfen.

Montag: Nimm deine nächste KI-Agenda und geh die fünf Fragen durch. Für den einen Block, in dem etwas entschieden wird, schreib das „Wann nicht einsetzen“ der Methode selbst in die Moderationsnotizen.



6. Wie sind wir vorgegangen, und wo liegen die Grenzen?

Zu den Dateien, die wir mit diesem Report veröffentlichen, gehören das Bewertungsraster, die Briefings und die Arbeitsanweisung. Alles, was in diesem Kapitel steht, lässt sich damit nachprüfen.

Die Befragung der Assistenten

Wir haben 40 Fragen jeweils auf Deutsch und Englisch an drei Assistenten gestellt, und zwar jede dreimal, weil KI-Antworten von Versuch zu Versuch schwanken. Claude und ChatGPT liefen mit eingeschalteter Websuche, Perplexity sucht ohnehin immer im Netz. Gefragt haben wir über die Programmierschnittstelle (API), damit jede Frage mit demselben Modell und denselben Einstellungen lief: Claude Opus 5, GPT-5 und Perplexity Sonar Pro. Die Chat-Apps können neuere Modelle wie Claude Fable 5.1 oder GPT-6 verwenden und andere Quellen heranziehen. Wer dort fragt, bekommt unter Umständen eine andere Antwort.

720

Antworten von KI-Assistenten, aufgezeichnet, ausgewertet und mit diesem Report veröffentlicht

Stellten wir dieselbe Frage dreimal in derselben Sprache, nannte ein Assistent in 54 von 60 Fällen dasselbe Werkzeug zuerst. Werkzeuge und Methoden haben wir anhand einer veröffentlichten Namensliste gezählt. Jede zitierte Website haben wir nach ihrem Betreiber eingeordnet, bekannte Seiten nach festen Regeln und seltene von Hand in einer Tabelle, die wir ebenfalls veröffentlichen. 5,4 Prozent der Links ließen sich nicht zuordnen und bleiben außen vor. „Zuerst genanntes Werkzeug“ ist eine Zählregel, keine Deutung der Empfehlung.

Der Agenda-Test

Jedes der drei Briefings ging einmal an die drei Assistenten, ohne Websuche und mit einer schriftlichen Arbeitsanweisung, die im veröffentlichten Bewertungsraster vollständig abgedruckt ist. Bewertet wurde nach acht Kriterien mit jeweils 0, 1 oder 2 Punkten. Was jede Punktzahl bedeutet, stand vor dem ersten Durchlauf fest:

#	Kriterium	2 Punkte bedeuten
1	Zeitrechnung	jeder Block hat eine Zeitangabe, die Summe stimmt, Pausen sind eingerechnet
2	Herkunft der Methoden	erkennbare, korrekt benannte Methoden mit Herkunft oder Quelle
3	„Wann nicht einsetzen“ beachtet	jede Methode passt zu Gruppe, Format und Ziel, oder der Widerspruch wird benannt
4	Passung zur Gruppengröße	die Gruppengröße wird in jedem Block ausdrücklich behandelt
5	Kleingruppen-Logistik (ein Briefing)	Gruppen, Aufteilung, Timing, Rückmeldung und Material sind genannt
6	Verständliche Sprache für Teilnehmende	Teilnehmende verstehen ihren Teil ohne Moderationsjargon
7	Treue zum Briefing	jede feste Vorgabe steht genau so in der Agenda wie im Briefing
8	Rückfragen (vages Briefing)	stellt die entscheidenden Fragen oder nennt Annahmen und bietet Varianten an

Kriterium 3 haben wir anhand der Hinweise „Wann nicht einsetzen“ aus unserer eigenen Methodensammlung geprüft. Die Warnungen stammen also von uns. Bewertet hat Claude, ohne zu wissen, welche Agenda von welchem Assistenten stammt; aufgelöst haben wir das erst danach. Eine zweite, unabhängige Claude-Instanz hat anschließend zwei Agenden je Assistent bewertet, ausschließlich mit dem Raster. Beide vergaben in 65 Prozent der 40 Fälle dieselbe Punktzahl und blieben damit unter den 70 Prozent, die wir als Mindestmaß festgelegt hatten. Jede der 14 Abweichungen haben wir geklärt, indem wir das Raster noch einmal gelesen haben; einen Mittelwert haben wir nie gebildet. Jedes Mal übernahmen wir die zweite Bewertung, 8-mal lag sie niedriger, 6-mal höher, und legten das Raster auch bei den Agenden so aus, die die zweite Instanz nicht gesehen hatte. Eine Gesamtpunktzahl gibt es bewusst nicht.

Die Auswertung der Forschung

Wir haben die Datenbanken OpenAlex und Semantic Scholar nach Studien seit 2022 durchsucht, die sich mit generativer KI in moderierten Gruppen, Meetings, Beteiligungsverfahren und Ideenfindung in Gruppen befassen. Das Suchprotokoll stand vor der ersten Anfrage fest. Anschließend haben wir in zwei Runden verfolgt, welche Arbeiten die gefundenen Studien zitieren und von welchen sie zitiert werden. Von 7.185 Treffern wurden 5.590 gesichtet und 139 aufgenommen. Die Praxisumfragen fanden wir über 40 Websuchen auf Deutsch und Englisch. Semantic Scholar hat unseren Zugriff begrenzt, sodass dort nicht jede Anfrage lief; das Suchprotokoll weist jede einzelne aus. Die typische aufgenommene Studie ist ein kurzes Online-Experiment mit bezahlten Teilnehmenden und kein echter Workshop.

Interessenkonflikt und Grenzen

Unser eigenes Produkt, Workshop Weaver, ist nicht Teil des veröffentlichten Agenda-Tests, weil wir es nicht glaubwürdig bewerten können. Produkte von Wettbewerbern haben wir ebenfalls nicht getestet. Ob irgendein Produkt die beschriebenen Lücken schließt, kann diese Ausgabe deshalb nicht sagen. Drei Briefings und ein einziger Testtag sind eine kleine Stichprobe. Der Test zeigt, was die Agenden von sich aus enthielten, und nicht, wozu die Assistenten auf Nachfrage in der Lage wären. Kein Mensch hat die Agenden bewertet; bewertet hat eines der getesteten Systeme selbst. Beide Untersuchungen liefen mit Modellen, die zum Zeitpunkt des Tests nicht mehr die neuesten waren.

Weiterverwendung und die nächste Ausgabe

Text und Daten stehen unter der Lizenz CC BY 4.0 und dürfen frei verwendet werden, solange der Report als Quelle genannt wird. Die Downloads zu diesem Report enthalten jede Antwort aus der Befragung samt Frage und zitierten Quellen sowie die Auswertungstabellen. Hinzu kommen die neun Test-Agenden, das Bewertungsraster und beide Bewertungen. Wer unsere Agenden mit unserem Raster bewertet und zu einem anderen Ergebnis kommt, sollte uns das mitteilen. Die nächste Ausgabe druckt es ab.

Fragen, die Facilitator:innen stellen

Ist dieser Report unabhängig?

Nein. Wir entwickeln Workshop Weaver, ein Werkzeug für die Workshop-Planung. Es ist nicht Teil des veröffentlichten Agenda-Tests, weil wir unser eigenes Produkt nicht glaubwürdig bewerten können. Andere Workshop-Werkzeuge haben wir ebenfalls nicht getestet. Wir veröffentlichen alle Antworten, das Bewertungsraster und beide Bewertungen, damit du unsere Arbeit nachprüfen kannst. Kapitel 2 zeigt, wie oft die Assistenten uns nennen.

Warum keine eigene Umfrage unter Facilitator:innen?

Zwei Umfragen gibt es bereits: den State of Facilitation von SessionLab (714 Antworten 2025/26) und die Umfrage von Leadership Strategies aus dem Jahr 2026, die die Zahl der Befragten nicht nennt. Eine dritte, kleinere Umfrage hätte wenig hinzugefügt. Wir haben stattdessen gemessen, was beide nicht erfassen: was die Assistenten empfehlen und was in ihren Agenden fehlt.

Welche Assistenten habt ihr getestet?

Claude, ChatGPT und Perplexity. In der Befragung war die Websuche eingeschaltet, im Agenda-Test ausgeschaltet. Kapitel 6 nennt die Modelle.

Darf ich die Daten weiterverwenden?

Ja. Alle Antworten, die Auswertungstabellen, die Test-Agenden und die Bewertungen stehen unter der Lizenz CC BY 4.0. Du kannst sie frei nutzen, solange du den Report als Quelle nennst.

Ist die deutsche Ausgabe eine Übersetzung?

Nein. Sie wurde auf Deutsch geschrieben. Die deutschen Fragen und das deutsche Briefing stehen im Vordergrund, und Kapitel 4 zeigt die deutsche Agenda im Original.

Gibt es eine Ausgabe 2027?

Das hängt davon ab, ob diese Ausgabe genutzt wird. Wenn ja, wiederholt die Ausgabe 2027 beide Untersuchungen mit den dann aktuellen Modellen und druckt jede Neubewertung ab, die uns Leser:innen schicken.

Zitieren als

Marian Kaufmann (2026). *State of AI Facilitation 2026*. Workshop Weaver. workshopweaver.com/state-of-ai-facilitation-2026. Text und Daten frei nutzbar mit Quellenangabe (Lizenz CC BY 4.0).