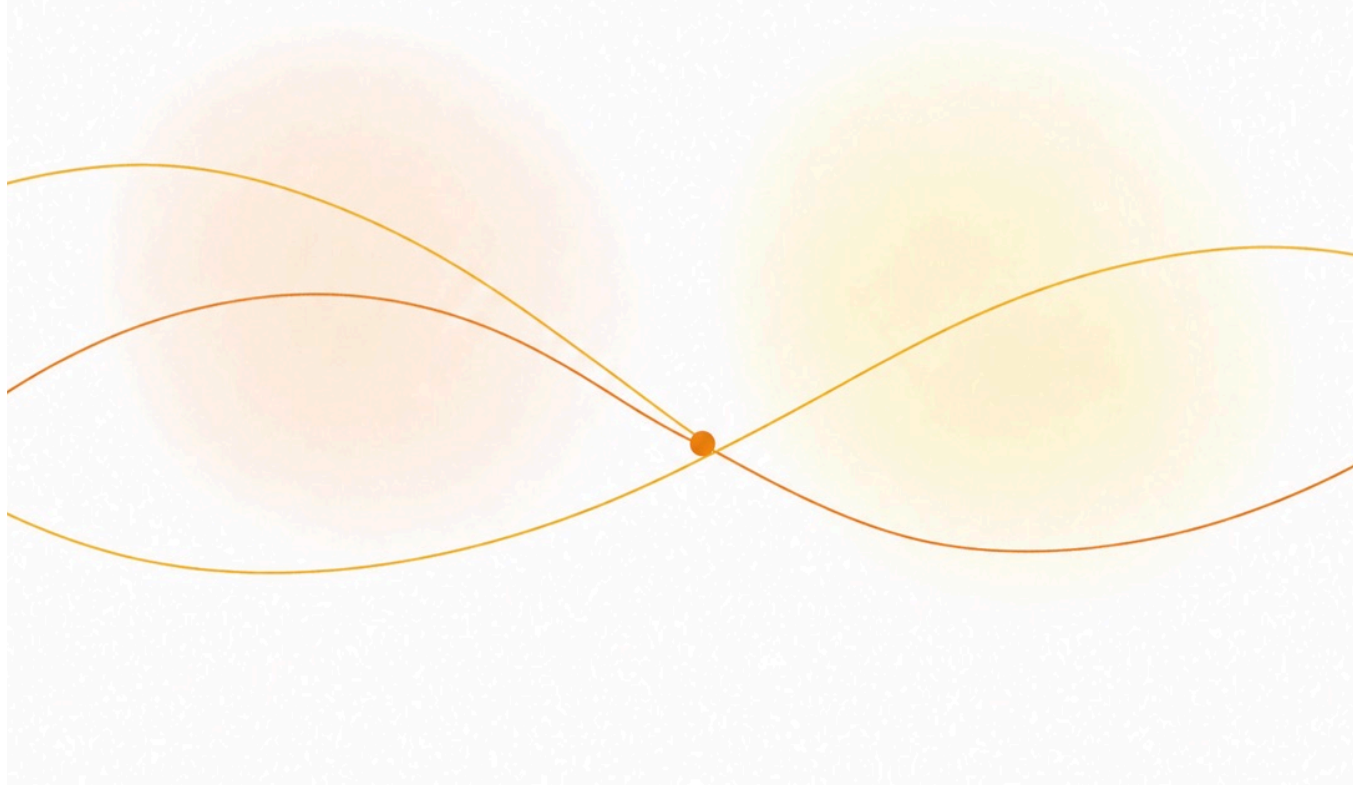


WORKSHOP WEAVER RESEARCH

State of AI Facilitation 2026

What AI assistants recommend
to facilitators, and what their
workshop agendas leave
unwritten

Nine AI-drafted workshop agendas, 720 assistant answers: the
timings hold, the method warnings go unwritten. What a facilitator
checks on Monday.



1. What did we find?

In the spring of 2026, the facilitators we work with stopped asking whether they should use AI to plan a workshop and started asking what they could trust it with. We make Workshop Weaver, a workshop-planning tool, so we had opinions and no data. This page is the data: 720 answers from three AI assistants to questions a facilitator would ask, and 9 agendas they drafted, scored against a sheet of what a facilitator checks before walking into the room. Read it as an early signal from a small sample. Every number carries its base and resolves to a published file.

- **In our agenda test, 8 of 9 agendas added up to the workshop length, breaks included, and the same eight kept every fixed point of the brief.**

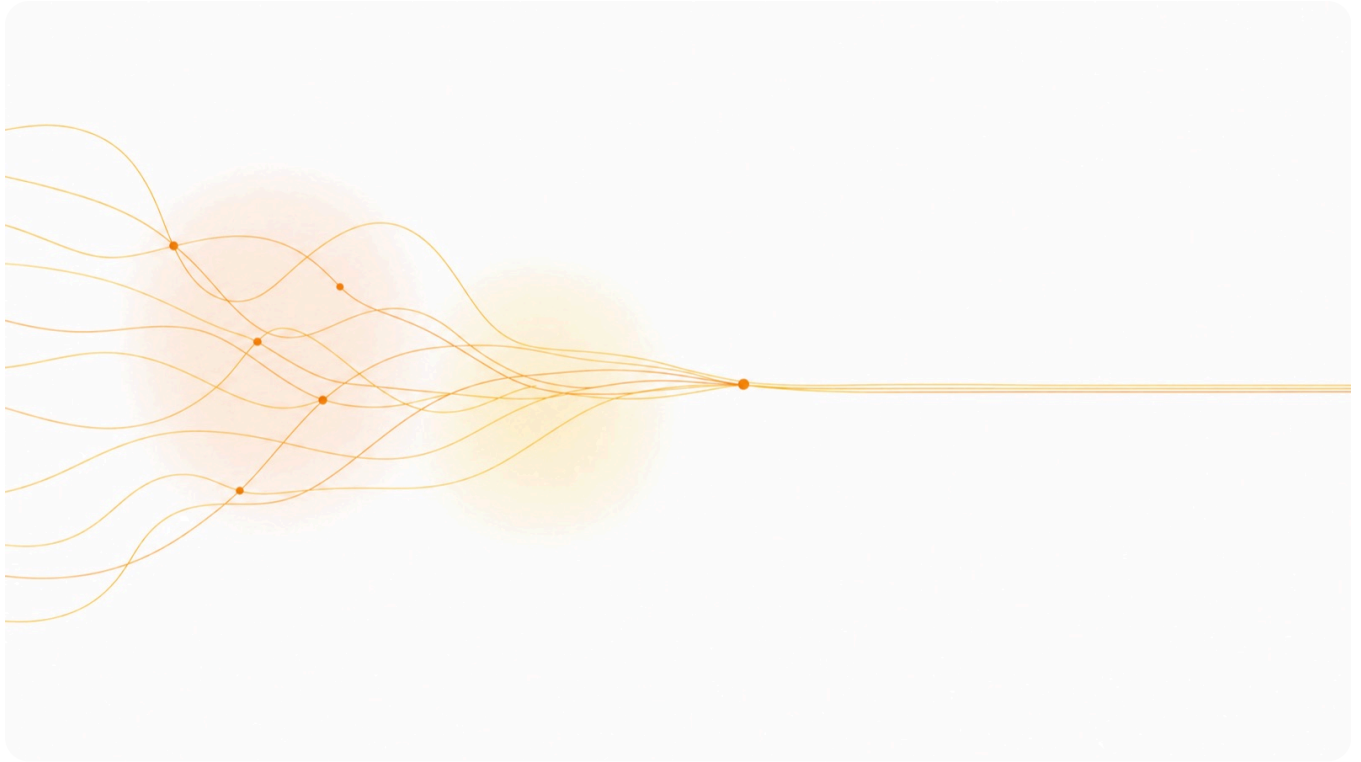
- **0 of 9 agendas wrote down where a method comes from. All nine named recognisable methods.**

- **7 of 9 agendas used a method in a setting its write-up warns against and did not say so. Five times it was a dot vote shaping a leadership decision.**

- **Of 108 agenda requests in our survey, 1 answer asked a question before it planned. On the deliberately vague request, 1 of 18 did.**

- **Six methods carry half of 541 method mentions, and Dot Voting appears in 114 of 270 method and agenda answers.**

Chapters 2 and 3 hold the data behind each line, chapter 4 one agenda block where the gap is visible on paper, chapter 5 the five questions we now put to every AI-drafted agenda. Chapter 6 says how we did it and where it falls short, starting with the fact that an AI did the scoring.



2. What do the assistants recommend when a facilitator asks?

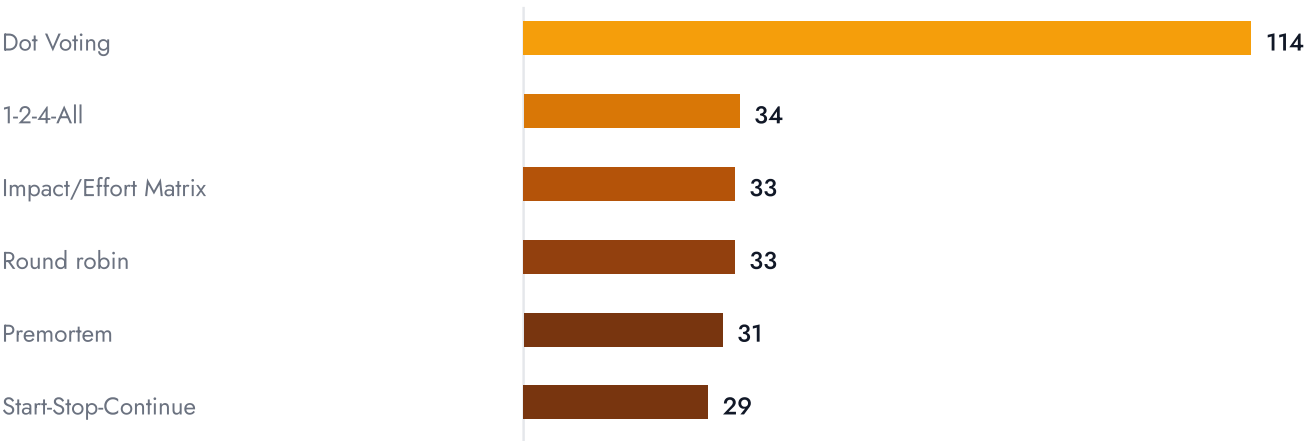
The same six methods for most questions, and a plan before a question. We put 40 questions to Claude, ChatGPT and Perplexity, in English and in German, three times each, with web search on. That gave 720 answers. We counted the tools and methods named and sorted every cited page by the kind of site behind it, by rules we publish with the raw data.

Methods: the same six names

Six methods account for half of all 541 method mentions: Dot Voting, 1-2-4-All, the Impact/Effort matrix, round robins, the premortem and Start-Stop-Continue. Dot Voting alone appears in 114 of 270 method and agenda answers. That is a sound toolkit. One check came out better than we expected. Where a question stated a group size, 4 of 144 answers recommended a method outside the size range in our library.

When a method fails gets less room in the answers. Asked when not to use Dot Voting, 5 of 18 answers mentioned group size, the first setting our write-up lists. In 28 of 72 answers to our strategy questions, Dot Voting or a premortem was recommended for a workshop with binding decisions on the table, the setting both write-ups warn against.

Exhibit 1 · Six methods make up half of all method mentions



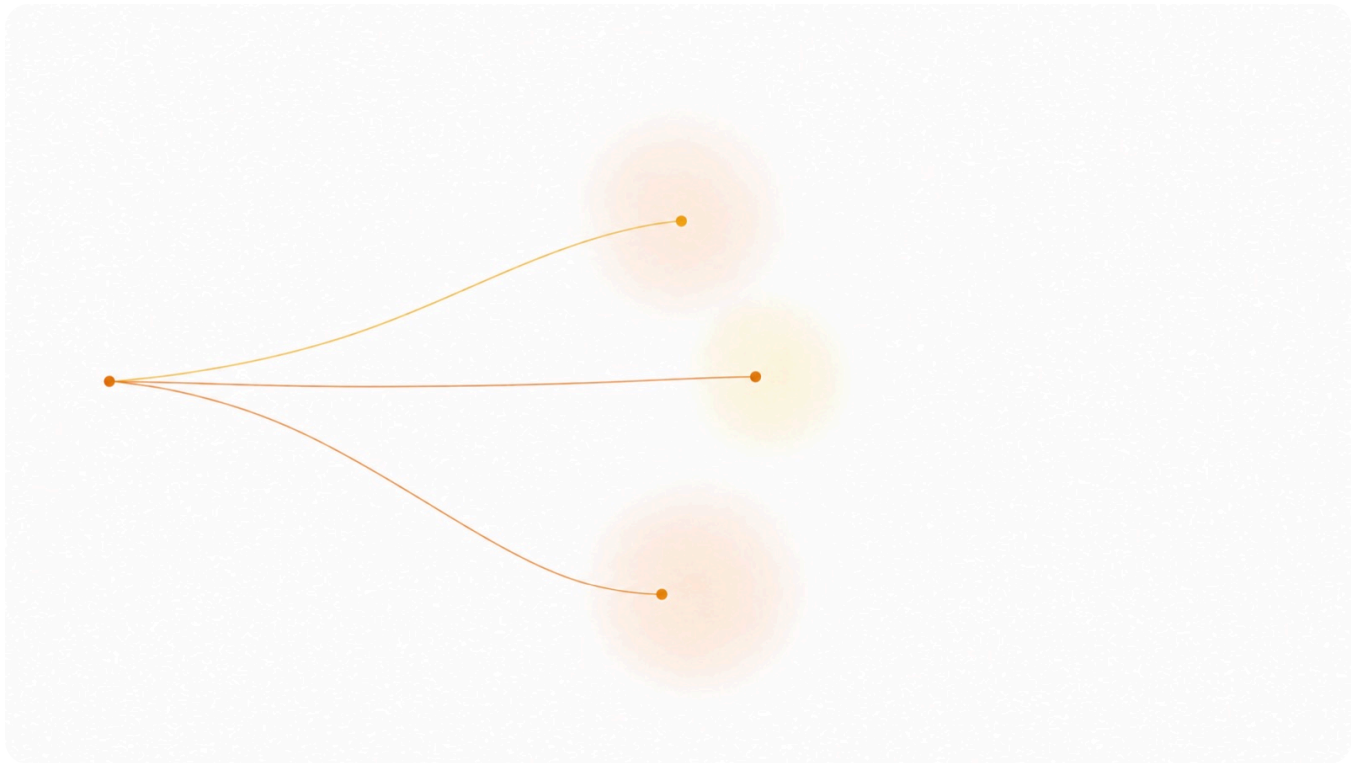
Source: our survey of AI assistants, September 2026. 270 answers to our questions about methods and agendas. Counted: answers that name the method.

Agendas: one question in 108

Of 108 agenda answers, 1 held back the plan and asked a question first. On the deliberately vague request, "I need a half-day retrospective for my team. We had a rough quarter.", 1 of 18 answers asked before planning. The rest produced a full agenda for a team whose size and format they had invented. Listing the assumption, as several did, is the honest version of the same move. Asking is the other.

Tools and sources: the short version

Asked which tool to use, the assistants named one built for workshop planning in 127 of 180 answers. The first name was SessionLab in 83 of 180; Workshop Weaver was named in 27 of 180. Of the 6,349 cited pages we could sort, 49.7 per cent belong to a company selling a tool or a service, and 14.7 per cent to a facilitator's own site. German questions got fewer local names: 62 of 90 German tool answers named nothing from the German-speaking market. One habit follows for a facilitator: check who owns the page the assistant cites.



3. How well do the assistants plan a workshop?

In our test the timings held, the brief was kept, and the method warnings went unwritten. We wrote 3 briefs: one fully specified in English, one in German with a fixed lunch break and a guest slot, and one deliberately vague. Each went once to Claude, ChatGPT and Perplexity with web search off, which gave 9 agendas. The instruction asked for a complete, timed agenda with a method for every block, what each side does, and materials. It did not ask for sources, for warnings about methods, or for questions, so the results show what these nine agendas contained by default.

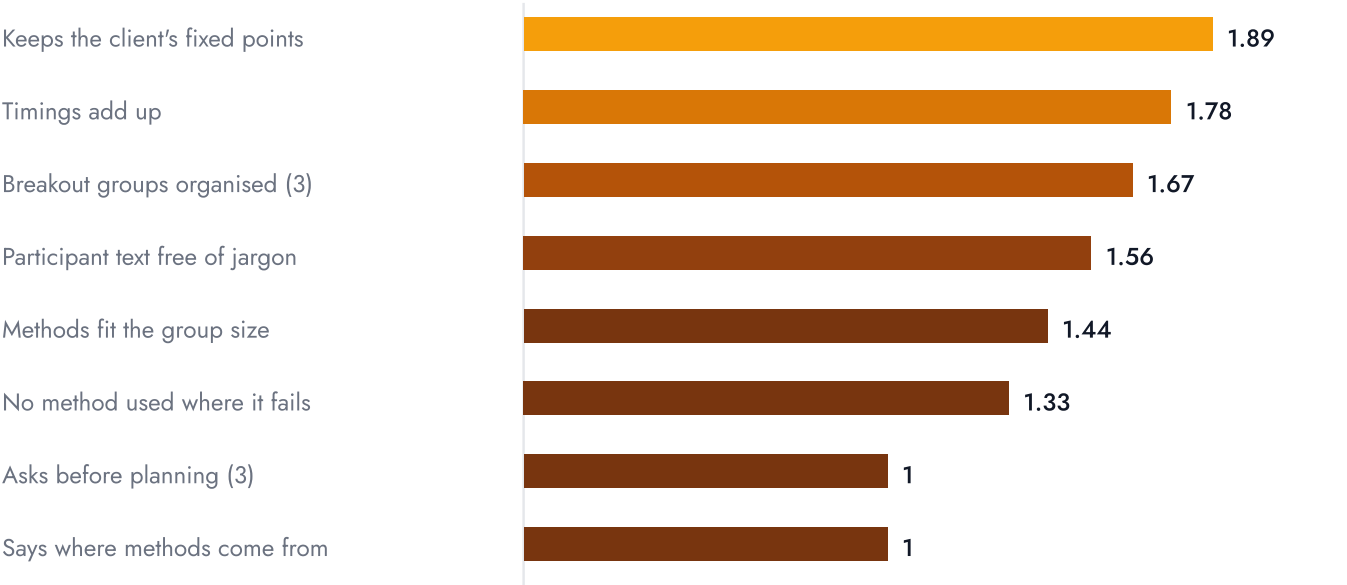
An AI rater, Claude, scored the agendas blind against a written sheet: eight criteria from 0 (missing) to 2 (fully met). Claude is also one of the assistants tested, and no person scored. A second rating agreed with the first in 65 per cent of 40 scores, under the 70 we had set. Read the scores as one machine reading against a fixed sheet; chapter 6 has the rest.

What the agendas got right

Timing and the brief. 8 of 9 agendas added up to the workshop length, breaks included and every block timed. The same eight kept every fixed element, in the German brief the lunch at half past twelve, the guest slot with no discussion, the end on the hour, the formal form of address. The ninth, Perplexity's answer to the vague brief,

announced a four-hour workshop and planned three. Where the brief asked for breakouts, 2 of 3 agendas stated the groups, the split rule, the timing and the report-back. Across all nine, 29 of 60 scores were full marks and 1 of 60 was a zero.

Exhibit 2 · The agendas keep the brief and the timings; they rarely say where a method comes from or when it fails



Source: our agenda test, September 2026. Nine agendas (three workshop briefs, each given to Claude, ChatGPT and Perplexity), scored 0 (missing) to 2 (fully done) per criterion; average shown. Breakout groups scored only on the brief that asked for them, asking first only on the vague brief (three agendas each).

What the agendas did not write down

Where a method comes from, and when it is a poor fit. 0 of 9 agendas said where a method comes from. Methods were named but not attributed or sourced, which is 1 point of 2 on our sheet for every agenda. A facilitator can check the method only against the assistant's own description of it.

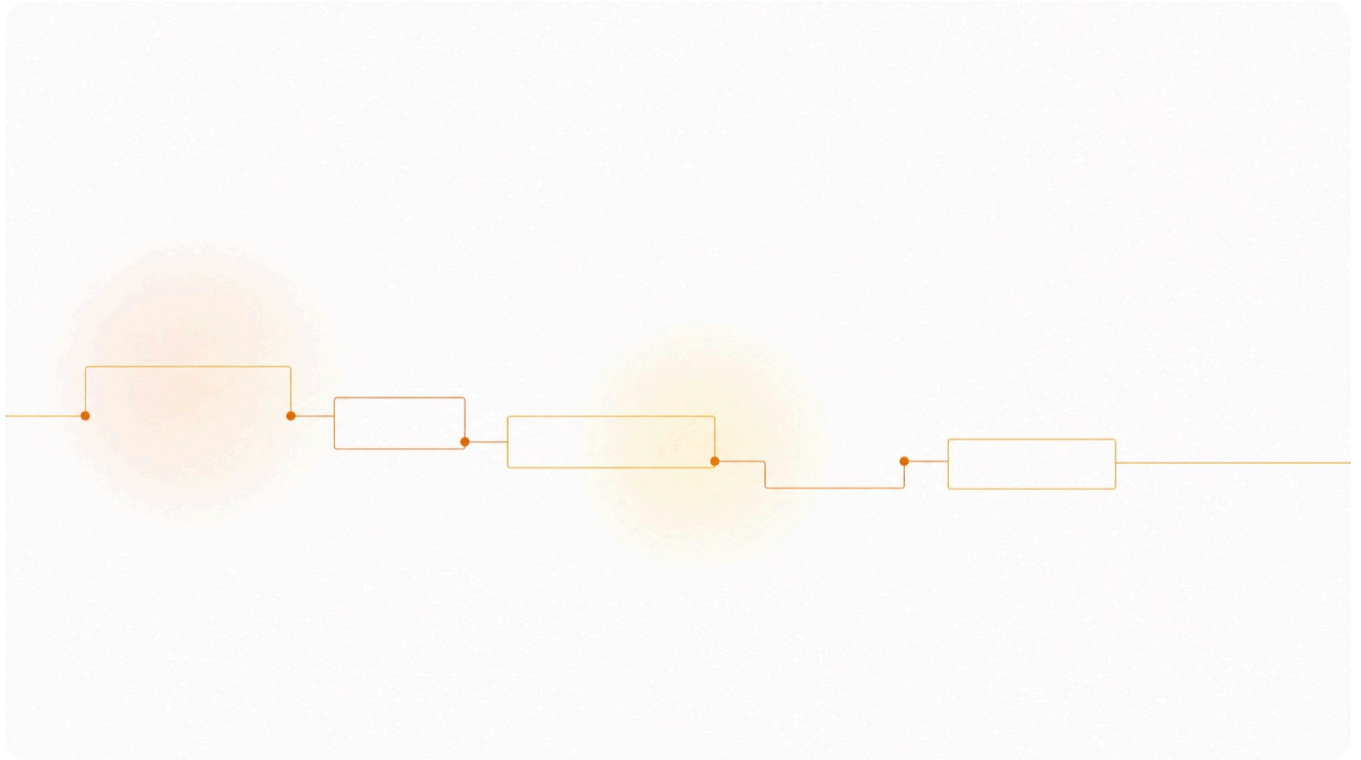
0
of 9 test agendas say
where a method comes
from

Warnings were the second gap. In 7 of 9 agendas, a method was used in a setting its write-up in our library warns against, without a word about it. Five times it was a dot vote shaping a binding decision with the leadership in the room. Twice it was a clustering run faster than its write-up allows, once alongside a one-word check-in round. Group size was stated for every block in 4 of 9 agendas. Participant text was cleanly separated from facilitator notes in 5 of 9.

The vague brief

On "I need a half-day retrospective for my team. We had a rough quarter", 0 of 3 assistants asked for clarification. Each stated an assumption in one line, four hours and a team of six to twelve, and planned. Our instruction invited exactly that, so the stronger evidence is the survey in chapter 2, where no instruction was given and 1 of 18 answers asked first. None of the three asked what a facilitator would ask first: what happened?

Whether a source, a warning or a question appears when asked for is something this test did not measure. The arithmetic held in these nine; the method choice is where a facilitator's reading time goes. The next chapter shows one block where that reading is needed.



4. What does a dot vote going wrong look like on paper?

Like a good agenda. The block below is from Claude's answer to the German strategy brief: ten people from a leadership team, one day, a three-year ambition and the first three initiatives. Apart from the provenance point that all nine missed, the agenda scored full marks on every criterion but one, and the gap is easiest to see where the planning is sound.

The block as the assistant wrote it

The agenda opens with an assumption, and the block at 15:15 picks the three initiatives. The quotes are our translation; the German original is in the German edition and in the published agendas.

The managing director is one of the ten participants and stays all day; ...

Then, at 15:15:

Two stages – (1) Impact/feasibility matrix: the clusters are placed together on a four-field matrix (15 min.); (2) dot vote: 3 sticky dots per person on the candidates in the "high impact" field (10 min.); (3) quick check of the top three for resistance (5 min.).

The fallback rules at the end of the agenda add one line, and it is the one to read twice:

If prioritising drags on: the dot vote decides, the matrix discussion is cut.

What the write-up warns about

Our Dot Voting write-up lists four settings in which not to use it, and two are on this page. The decision is binding and high-stakes, and dots measure preference, not feasibility or evidence. A strong hierarchy is in the room with votes in public, the setting in which the write-up asks for an anonymous format. The pitfalls add a third: dots placed in turn anchor later voters on earlier dots, and the block does not say whether they go up at once. The fallback rule turns the risk into the plan: under time pressure, the preference signal becomes the decision.

The fair reading

A facilitator who defends this block has a case. On the main path the dots only narrow a shortlist. The matrix before them handles feasibility, the alternative the write-up itself suggests, and the resistance check after them is where the decision is made. The opening block announces consent as the decision mode for the day. Our first rater gave the block full marks on that reading; the second read "borderline fit, no warning given" and scored it 1, and we sided with the second. Both readings share one fact: the condition and the fallback are on the page, and the warning is not.

The fix

Two sentences a facilitator would add to the block. "Dots go up at the same time, and the managing director votes last or not at all. The dot picture is a shortlist; the decision is the resistance check." Then one line swapped in the fallback. If time runs short, the matrix discussion is cut and the resistance check stays. The method can stay. The warning needs writing where the facilitator will read it at 15:14.

5. What do you check before you run an AI-drafted agenda?

Five things, one minute each, and the arithmetic is not one of them. In SessionLab's State of Facilitation 2026 (714 responses), 38.7 per cent of facilitators use generative AI often. Its 2025 edition (1,050 responses) put full-agenda design at 25.9 per cent of AI users, about one in four. Our search of the published research found 139 studies on AI in facilitated groups, all run on models from 2025 or earlier, so they say little about today's assistants and something about groups. In Alsobay et al. (2026), groups with an AI facilitator shared more of what they knew than groups with a human one and did not choose the best option more often. In Parisi et al. (2026), the AI facilitator moved the group's budget by restating outlier proposals. Our search found no study that compared an AI-drafted agenda with a facilitator's, and none that tested current models in facilitated groups. Every study is listed with its extraction notes in the published data.

The audit checklist

BEFORE YOU RUN AN AI-DRAFTED AGENDA

1. Does every block add up to the workshop length, breaks included, and does the agenda say so?
2. For each method, what group size was it written for, and what group do you have?
3. For each method, is there a "when not to use" in its literature, and does it apply?
4. Does the participant text stand on its own without the facilitator's notes and jargon?
5. Did the brief have fixed elements, and are they exactly where the client put them?

What this means for 2027

We expect general assistants to stay reliable on timing and fixed requirements, and method warnings and clarifying questions to stay absent unless the instruction asks for them. These are expectations from nine agendas and one day of runs, not findings. The 2027 edition will check both against the models of that year.

Monday: Take your next AI-drafted agenda and run the five questions on it. For the one block that decides something, write the method's "when not to use" into the facilitator notes yourself.

6. How did we do this, and what are the limits?

The files published with this page include the scoring sheet, the briefs and the instruction, so everything below can be checked.

The survey of AI assistants

40 questions, each in English and German, went to 3 assistants, 3 times each, because AI answers vary from one try to the next. Claude and ChatGPT ran with web search on; Perplexity always searches. We asked through each assistant's developer access (API), so every question ran with the same model and settings: Claude Opus 5, GPT-5 and Perplexity Sonar Pro. The chat apps may run newer models, such as Claude Fable 5.1 or GPT-6, and look up different pages. A facilitator typing the same question may see a different answer. Asked the same question in the same language three times, an assistant named the same tool first in 54 of 60 cases. Tools and methods were counted against a published list of names. Each cited site was sorted by type, well-known ones by rule and rarer ones by hand in a table we publish. 5.4 per cent of links could not be sorted and are left out. "Tool named first" is a counting rule, not a reading of the advice.

720

assistant answers recorded, sorted and published with this report

720

assistant answers
recorded, sorted and
published with this
report

The agenda test

3 briefs went once to each of the three assistants with web search off, under one written instruction, quoted in full in the published scoring sheet. Eight criteria, each scored 0, 1 or 2 against definitions fixed before the first run:

#	Criterion	2 means
1	Timing arithmetic	every block timed, sums exactly, breaks included
2	Method provenance	recognisable methods, correctly named, with an origin or a source
3	"When not to use" respected	every method fits the group, format and goal, or the mismatch is flagged
4	Group-size fit	group size handled explicitly in each block
5	Breakout logistics (one brief)	groups, split rule, timing, report-back and materials all stated
6	Participant-safe language	a participant could read their part without facilitator jargon
7	Brief fidelity	every fixed element exactly as briefed
8	Clarifying behaviour (vague brief)	asks the questions that matter, or states assumptions and offers variants

Criterion 3 was checked against the "when not to use" notes in our own method library, so the warnings are ours. The rater was Claude, scoring blind; we unmasked the agendas afterwards. A second, separate Claude instance scored two agendas per assistant with only the sheet. The two agreed in 65 per cent of 40 scores, below the 70 per cent we had set as the minimum. Each of the 14 disagreements was settled by re-reading the sheet, never by averaging. The second rating was adopted every time, lower in 8 cases and higher in 6, and the same reading was applied to the agendas it had not seen. No overall score exists, on purpose.

The research review

We searched OpenAlex and Semantic Scholar for studies since 2022 on generative AI in facilitated groups, meetings, deliberation and group ideation, following a protocol written before the first query. We then tracked references and citations for two rounds. Of 7,185 records, 5,590 were screened and 139 included. 40 web searches in English and German found the practitioner surveys. Semantic Scholar limited our access, so some queries did not run there; the search log marks which. The typical included study is a short online experiment with paid participants, not a workshop.

Conflicts and limits

We make Workshop Weaver, a workshop-planning tool. It is not part of the published agenda test, because we cannot score our own product credibly, and no competitor's product is in it either. So this edition does not say whether any product closes the gaps it describes. Three briefs and one day of runs are a small sample, and the test shows what the agendas contained by default, not what the assistants can do when asked. No person scored the agendas, and the rater is one of the systems tested. Both studies ran on models that were not the newest available at the time.

Reuse and the next edition

Text and data are free to reuse under the CC BY 4.0 licence, as long as you credit the report. The downloads next to this page hold every survey answer with its question and cited pages, the summary tables, the nine agendas, the scoring sheet and both sets of scores. If you score our agendas with our sheet and reach a different result, tell us, and the next edition will print it.

Questions facilitators ask

Is this report independent?

No. We make Workshop Weaver, a workshop-planning tool. It is not part of the published agenda test, because we cannot credibly score our own product, and no other workshop tool is in it either. We publish every answer, the scoring sheet and both sets of scores so anyone can check our work, and chapter 2 reports how often the assistants name us.

Why no survey of facilitators?

Two surveys already exist: SessionLab's State of Facilitation (714 responses in 2025/26) and Leadership Strategies' 2026 survey. A third, smaller one would add noise. We measured what neither of them does: what the assistants recommend, and what their agendas leave unwritten.

Which assistants did you test?

Claude, ChatGPT and Perplexity: in the survey with web search switched on, in the agenda test with web search switched off. Chapter 6 names the models.

Can I reuse the data?

Yes. All answers, the summary tables, the test agendas and the scores are published under CC BY 4.0: free to reuse, as long as you credit the report.

Is there a German edition?

Yes. It was written in German rather than translated: the German questions and the German brief come first.

Will there be a 2027 edition?

That depends on whether this one gets used. If it does, the 2027 edition repeats both studies with the models of the day and prints any re-scoring readers send us.

Cite as

Marian Kaufmann (2026). *State of AI Facilitation 2026*. Workshop Weaver. workshopweaver.com/state-of-ai-facilitation-2026. Text and data free to reuse with credit (CC BY 4.0 licence).