Frontier Fit

Facilitator guide, production SOP, and prompt library

Independent candidate prototype prepared by Amir Daniel | August 22, 2026

Independent candidate prototype created using publicly available HBS research. Not an official Harvard Business School product. Requires faculty/SME validation before delivery.

Session purpose

Help executives and managers make an evidence-based decision about where AI should and should not hold authority inside one real workflow.

Audience

Executives, functional leaders, founders, operating managers, and risk/enablement leaders. No engineering background required.

Prerequisites

Participants bring one recurring workflow they understand. They do not need an AI account. The facilitator owns the demonstration and provides a fictional example for anyone who cannot use real work.

Learning objectives

Participants will be able to:

1. Decompose one workflow into 4-7 observable tasks.
2. State an evidence-based hypothesis about AI fit and uncertainty.
3. Design a representative comparison with a baseline, rubric, edge case, and stop rule.
4. Assign AI and human authority with an accountable checkpoint.
5. launch a seven-day experiment with a go/revise/stop decision.

Preparation

One week before

- Confirm participant roles, delivery mode, tools, data policy, and accommodations.
- Obtain faculty/SME approval of the research brief and all study claims.
- Confirm which AI environment, if any, is approved for the demo.
- Send participant note: bring one recurring workflow; do not bring confidential data.
- Load deck, workbook, prepared demo outputs, timer, polling method, and offline copies.
- Re-check source links and study figures.

48 hours before

- Print or distribute workbook.
- Test captions, microphone, display, links, and screen share.
- Place prepared AI outputs in local files, not only a web tab.
- Confirm room tables support pairs; for virtual, create pairs or small breakout rooms.
- Review likely misconceptions and difficult questions.

30 minutes before

- Open slide deck in presenter mode.
- Open the demo input, first output, corrected output, and evidence note.
- Put the 60-minute run-of-show and cut options where the facilitator can see them.
- Confirm offline timing and the contact for technical support.

Materials and setup

In person

Screen, presenter computer, visible timer, microphones for large rooms, printed workbook, pens, optional sticky notes. Seat participants in pairs across functions where possible. Provide a quiet location and digital workbook for anyone who needs an alternative to handwriting.

Virtual

Captions on; chat enabled; breakout pairs preassigned; workbook link posted before the session; polling accessible by chat as an alternative. Ask participants to keep cameras optional. Provide a dial-in or low-bandwidth path.

Technology

Live AI is optional. Never rely on it for the learning outcome. If used, enter only fictional data. Prepared outputs are the source of truth for timing.

Minute-by-minute guide

Time	Segment	Objective	Asset	Cut option
0-5	Provocation and vote	Expose prior belief	Slides 1-3	Take two shares instead of four
5-12	Research insight	Accurate mental model	Slides 4-10	Combine study design and limits in one explanation
12-20	Frontier Fit Loop	Framework comprehension	Slides 11-13	Use one example, not two
20-30	Live demo	Experience verification failure	Slides 14-16	Use prepared output only
30-45	Frontier Fit Map	Apply to real work	Slides 17-20	Complete one task rather than full workflow
45-53	Peer challenge and debrief	Revise judgment	Slides 21-22	One partner question and one plenary share
53-58	Seven-day experiment	Plan transfer	Slide 23	Complete owner/metric/stop/review only
58-60	Commitment and post-check	Demonstrate next action	Slide 24	Written commitment only

Facilitation script

0-5 minutes - provocation

SAY

“This is not a session about whether AI is good or bad. It is about a harder operating question: where should AI hold authority inside a workflow when polished output is not proof of correctness?”

“Imagine a tool makes your team roughly 25 percent faster on one kind of work. On another task, people using it are materially less accurate. Would you deploy it?”

ASK

Vote: yes, no, or not enough evidence. What is the minimum evidence you need?

DO

Collect the vote before explaining the study. Ask two participants with different answers to share. Write their evidence requirements where everyone can see them.

WATCH FOR

Participants debating vendors or security before defining the task. Acknowledge the concern and return to the decision: “What task and consequence are you imagining?”

DEBRIEF

“The disagreement is useful. Most blanket AI policies hide different assumptions about task, evidence, and consequence.”

5-12 minutes - research insight

SAY

“The research team partnered with BCG on a preregistered randomized experiment with 758 consultants. After a baseline, participants received no AI, GPT-4, or GPT-4 plus a brief overview.”

“On 18 innovation and go-to-market subtasks designed within the tested model's frontier, AI users completed 12.2 percent more work, worked 25.1 percent faster, and produced work the HBS AI Institute's 2026 interpretation describes as about 32 percent higher in quality.”

“On a different brand-strategy task, the result reversed. The no-AI group was correct 84.5 percent of the time, compared with 70.6 percent for GPT-only and 60 percent for GPT plus the overview.”

“This does not mean prompt training lowers accuracy. It means that in this designed task, a capable and persuasive model could pull skilled people toward the wrong conclusion.”

ASK

What makes the outside-frontier failure more dangerous than a visibly broken answer?

EXPECTED RESPONSES

Confidence; speed; managerial trust; lack of independent verification; coherent narrative; hidden qualitative evidence; error propagates downstream.

WATCH FOR

- “AI is inaccurate.” Correct: performance varied by task and configuration.
- “The frontier is known.” Correct: tasks were pretested in the study; organizations normally must learn through evidence.
- “This applies to every worker.” Correct: the sample was highly selected consultants.

DEBRIEF

“The unit of analysis is the task inside a workflow, and the frontier moves. The study gives us a problem to manage, not a universal task list.”

Source: primary paper; HBS AI Institute summary.

12-20 minutes - framework

SAY

“I translated the research into a candidate tool called the Frontier Fit Loop: Decompose, Predict, Test, Assign, Monitor. This is not a framework named by the authors. It is a prototype translation I would validate with faculty.”

Walk through the five steps using the customer-escalation workflow.

ASK

Which step does your organization most often skip?

DO

Have participants read the five step names to a partner and underline the step they skip. Then ask them to write one promising and one uncertain task from the workflow they brought.

WATCH FOR

The word “restricted” being interpreted as permanent prohibition. Explain that it means human-led by default under current evidence and consequence.

DEBRIEF

“Prediction is a hypothesis. Test and monitor keep the framework from becoming false certainty.”

20-30 minutes - live demonstration

SAY

“You are choosing a market for a six-week launch test. First, make your own decision from the metrics.”

Show Northstar and Harbor metrics. Give 45 seconds. Then run or reveal Prompt 1 and the prepared Northstar recommendation.

“Rate the recommendation's persuasiveness from one to five. Now rate the evidence completeness.”

Reveal the distribution-agreement note.

ASK

What changed: the market, the evidence, or the decision definition?

EXPECTED RESPONSES

Evidence and constraint; time horizon made the decision different; the first prompt omitted the most important context; the attractive market may remain a longer-term option.

DO

Run Prompt 2 or reveal the corrected prepared output. Ask participants to circle which framework step failed first.

WATCH FOR

“The prompt was bad” as the only diagnosis. Ask: “Who owns defining the decision and ensuring authoritative evidence is present? What would prevent this failure at scale?”

DEBRIEF

“Prompt quality matters, but the deeper control is workflow design: evidence inventory, representative tests, authority, and verification.”

30-45 minutes - participant exercise

SAY

“Choose one consequential, repeatable task. If your workflow is too broad, ask what observable action begins and ends.”

“You have six minutes to map the current workflow, four minutes to design the test, and three minutes to assign authority. I will call checkpoints.”

DO

Minute 33: scan for job-level labels. Minute 36: require a current baseline. Minute 39: require one edge case. Minute 42: require an independent evidence source and stop rule.

ASK

At each checkpoint: “What would a successful demo fail to reveal?”

WATCH FOR

Confidential data; high-stakes live experiments; speed-only metrics; generic “human review”; no named owner; test cases selected to make AI look good.

DEBRIEF

Do not debrief yet. Ask participants to star the weakest assumption for peer review.

45-53 minutes - peer challenge and debrief

SAY

“Your partner is not reviewing your prompt. They are reviewing the decision system.”

ASK

Where could a polished error survive? Is the test representative? Is the metric meaningful? Is the action reversible?

DO

Pairs take three minutes each. In plenary ask: “Where did you move the human checkpoint, and what evidence caused the move?”

EXPECTED RESPONSES

Earlier evidence validation; separate draft from approval; route edge cases; preserve independent judgment; add blind scoring; use business outcome instead of volume.

DEBRIEF

“The research pattern is the lesson: same professionals, same model, different task fit. Your workflow needs differentiated authority.”

53-58 minutes - apply

SAY

“Make the first experiment small, reversible, and decision-producing. A pilot without a stop rule is a rollout in disguise.”

DO

Participants complete owner, first case, baseline, quality metric, stop rule, and review date. Ask them to create the calendar action now if appropriate.

WATCH FOR

Vague dates, no owner, launch defined as success, or metrics impossible to collect in seven days.

DEBRIEF

“The output is a go, revise, or stop decision - not proof that you used AI.”

58-60 minutes - commitment

SAY

“The frontier is not a line someone else can map once for your organization. It moves by task, context, model, and consequence. Your advantage is a disciplined way to test, assign judgment, and learn before the risk travels.”

ASK

Complete: “In the next seven days, I will test ___ before allowing AI to ___.”

DO

Collect the post-performance task and feedback link. Remind participants of the follow-up date.

Difficult questions

“Doesn't a newer model make this research obsolete?”

Suggested response: “The exact location of the frontier changes. That is a limitation and the reason the operating principle remains relevant: do not infer fit from a model name. Re-test the task, configuration, and consequence.”

“Can we use benchmarks instead of testing our workflow?”

“Benchmarks can inform a hypothesis. They rarely contain your inputs, exceptions, decision rights, data quality, and cost of error. Use them as prior evidence, not as the final acceptance test.”

“Why not just require human review everywhere?”

“A generic review can become ceremonial, expensive, or vulnerable to anchoring. Specify what the human checks, which evidence is independent, and what triggers escalation. The appropriate control should match consequence and observability.”

“Is AI-led ever responsible?”

“Potentially, for a bounded task with repeated strong evidence, easily detected and reversible errors, low consequence, sampled review, and monitoring. This session does not certify any specific task.”

“What about confidential data?”

“Use only approved enterprise environments and follow organizational policy. For this session, abstract or fictionalize the workflow; do not enter sensitive data.”

“What if the AI refuses or produces a different live answer?”

“That variability is part of the lesson. Use the prepared outputs to protect the learning objective, then note that configuration and nondeterminism are re-test triggers.”

Accessibility and inclusion during delivery

- State instructions verbally and in writing before timed work.
- Use captions and a microphone; repeat participant questions.
- Never rely on color alone. Read chart labels and figures aloud.
- Allow typed, written, spoken, or fictional-workflow participation.
- Give an example before asking for independent application.
- Avoid cold calling after disclosure of uncertainty; invite pass.
- Pair across function, not by presumed fluency.
- Pause after each research number and explain what it means.
- Provide large-print and accessible digital workbook.
- Keep motion optional and avoid rapid animations.

Technology contingency plans

Live AI unavailable

Reveal prepared Prompt 1 output and corrected output. Ask participants what they would test next. Learning impact: none if timing is preserved.

Polling unavailable

Use fingers 1-3, colored cards with text labels, or chat A/B/C.

Breakouts unavailable

Use self-review with the four peer-challenge questions, then ask for two volunteer examples.

Workbook link unavailable

Show the six minimum fields on Slide 23: owner, case, baseline, quality metric, stop rule, review date. Send full workbook after.

Session shortened to 45 minutes

Use prepared demo; teach research in five minutes; complete one selected task; peer challenge one question; require a five-field action card.

Session extended to 90 minutes

Add a second test case, blind scoring calibration, cross-functional challenge, and operating-mode redesign across the full workflow.

Assessment administration

Before

Administer the two scenario questions and a five-minute baseline performance task. Do not teach answers first.

During

Record the hinge-question split, percentage with task-level maps by minute 36, percentage with edge cases by minute 39, and percentage with true stop rules by minute 42.

After

Score the final plan from 0-3 on decomposition, frontier hypothesis, test design, authority/guardrail, and metrics/action. Use two raters for the first pilot and discuss discrepancies greater than one point.

Later

Seven days: test initiated and evidence captured. Thirty days: pilots advanced/revised/stopped and metrics versus baseline. Ninety days: sustained outcome, exception/override pattern, and transfer to a new workflow.

Feedback instrument

Ask only:

1. Which step can you now perform more reliably? Point to evidence in your worksheet.
2. Where were you least sure what to do next?
3. Which research claim or limitation remained unclear?
4. Which part needed different timing?
5. Did your workflow produce a concrete next action? If not, what blocked it?
6. What example, phrase, visual, or technology requirement excluded or slowed you?
7. What one change would improve your ability to run the seven-day test?

Do not use satisfaction as proof of learning.

SOP - AI-Assisted Experiential Learning Production

1. Intake

- [ ] Audience, context, duration, modality, sponsor, deadline confirmed
- [ ] Source of record, version, publication status, authors, rights logged
- [ ] Required asset list and channels confirmed
- [ ] Data classification and tool policy confirmed
- [ ] Project ID, owner, decision log, folder, and v0.1 created
- [ ] Faculty/SME review dates scheduled

Exit: signed intake brief and source manifest.

2. Research validation

- [ ] Full primary source read
- [ ] Question, method, sample, measures, findings, nulls, limitations extracted
- [ ] Claim-evidence-limit ledger built with page references
- [ ] Direct findings separated from author interpretation and prototype implications
- [ ] Every high-consequence claim manually verified
- [ ] Missing public evidence marked “requires faculty validation”

Exit: faculty-ready research brief.

3. SME collaboration

- [ ] Brief sent before review
- [ ] Interpretation corrections captured
- [ ] Non-negotiable nuance and prohibited extrapolations recorded
- [ ] Executive decision/relevance question agreed
- [ ] Approval status and owner logged

Exit: approved interpretation or explicit prototype-only status.

4. Objectives and design

- [ ] 3-5 observable objectives
- [ ] Knowledge, skill, practice, assessment, and criterion aligned
- [ ] Real-work transfer task chosen
- [ ] Learn/See/Do/Reflect/Apply rhythm mapped
- [ ] Measurement defined before asset production
- [ ] Accessibility needs included in the architecture

Exit: approved alignment map.

5. Production

- [ ] Single controlled content source established

- [ ] Claim, objective, activity, and asset IDs assigned
- [ ] Source-bounded AI prompts used
- [ ] Low-fidelity narrative reviewed before polish
- [ ] Deck, workbook, guide, assessment, and media derived consistently
- [ ] Prototype and faculty-validated content visibly distinguished

Exit: complete alpha assets with no placeholders.

6. Review lanes

- [ ] Faculty/SME: claims, nuance, implications
- [ ] Instructional design: alignment, transfer, cognitive load
- [ ] Facilitator: timing, language, likely responses, fallbacks
- [ ] Accessibility: structure, contrast, alt text, captions, keyboard, alternatives
- [ ] Program/GTM: audience, logistics, claims, disclaimer

Exit: review matrix closed or remaining issue accepted by named owner.

7. QA and release

- [ ] Research claims match ledger and sources
- [ ] Source URLs and links pass
- [ ] Terminology and figures match across assets
- [ ] No facilitator answer exposed in participant materials
- [ ] Timing reconciles across deck and guide
- [ ] Every page and slide rendered and visually inspected
- [ ] Accessibility audit completed
- [ ] Virtual/in-person technology and offline fallback tested
- [ ] Release manifest and notes generated

Version: major for objective/research/experience redesign; minor for content/activity revision; patch for non-substantive production fix.

8. Delivery

- [ ] T-48 roster, accommodations, tools, files, facilitator roles
- [ ] T-30 room/virtual test and offline assets open
- [ ] During: timing, questions, misconceptions, deviations logged
- [ ] No unnecessary personal or confidential data collected

9. Assessment and iteration

- [ ] Performance, feedback, observation, and transfer evidence joined
- [ ] AI-assisted analysis checked against raw evidence
- [ ] Minority and accessibility signals reviewed
- [ ] Keep/change/test decision documented
- [ ] Faculty review triggered for research-interpretation changes
- [ ] Version, owner, risk, and validation measure recorded

10. Archive

- [ ] Source manifest and approved research brief
- [ ] Editable and final assets
- [ ] Release manifest and change log
- [ ] Assessment instruments and de-identified results
- [ ] Approval record, superseded labels, retirement date

AI production responsibility matrix

Stage	AI may accelerate	Humans remain responsible	Faculty approval
Ingestion	Parse files and metadata	Rights, version, privacy	As needed
Evidence	Draft claim ledger	Read and verify every material claim	Yes
Implications	Generate cautious options	Relevance and reasoning bridge	Yes
Objectives	Suggest measurable alternatives	Select outcomes and alignment	Review
Exercises	Generate fictional variants	Validity, realism, inclusion, safety	Review if research embedded
Assets	Draft transformations and alt text	Coherence, voice, layout, release	Substantive claims
Assessment	Draft items/rubrics	Construct validity and thresholds	Substantive items
QA	Compare assets and flag drift	Resolve issues and approve	Final claim review
Delivery	Prepare reminders and summaries	Facilitate and adapt	No
Revision	Cluster and draft change log	Inspect raw evidence and decide	Interpretation changes

Never automate final research interpretation, ethical approval, nuanced grading, faculty-attributed publication, or curriculum change without accountable human review.

Prompt library

Shared safeguards

Use only supplied sources. Do not invent findings, methods, quotations, page numbers, citations, or author intentions. Preserve uncertainty and limitations. Distinguish direct findings, author interpretation, your inference, and proposed instructional translation. Do not change correlation into causation or generalize beyond sample and task. If evidence is missing, write “not established in the supplied source.” Every research claim needs a source location. List unresolved issues under “Requires faculty validation.”

1. Research extraction

Act as a research analyst supporting an instructional designer. From the primary source, extract research question, context, sample, assignment, intervention, comparison, measures, analytical approach, preregistration/publication status, findings with effect sizes, null findings, limitations, and open questions. Create a claim-evidence-limit table. Apply Shared safeguards.

2. Evidence verification

Audit the claim ledger against the primary source. For each claim return supported, partially supported, unsupported, or overstated; exact source location; wording change; and faculty-validation need. Treat causal verbs, population claims, and numbers as high risk. Apply Shared safeguards.

3. Executive implications

Generate five cautious executive implications. For each show the supporting finding, reasoning bridge, boundary conditions, decision informed, and misinterpretation risk. Label proposed translations separately from author conclusions. Apply Shared safeguards.

4. Learner analysis

Create learner personas from supplied audience evidence: role, context, prior fluency, motivation, fear, misconception, task need, accessibility need, value signal, and disengagement risk. Do not fabricate demographic traits. Apply Shared safeguards.

5. Learning objectives

Using backward design, propose 3-5 observable objectives. For each include knowledge, skill, practice, assessment, success criterion, and supporting research claim. Reject vague verbs unless paired with performance. Apply Shared safeguards.

6. Workshop architecture

Design three 60-minute architectures using Learn, See, Do, Reflect, Apply. No passive block over ten minutes. Map every segment to objective and assessment. Include transitions, cognitive-load rationale, and technology-failure variant. Apply Shared safeguards.

7. Exercise generation

Create three real-work exercises requiring a decision or artifact. Include objective, instructions, inputs, outputs, time, rubric, facilitator role, common mistakes, accessibility alternative, and transfer step. Do not reenact or copy the study. Apply Shared safeguards.

8. Case/scenario generation

Create a fictional executive case with invented data and one decisive qualitative constraint conflicting with surface quantitative evidence. Label all entities fictional. Produce learner version, answer key, likely AI error, and alternate interpretations. Do not borrow proprietary case content. Apply Shared safeguards.

9. Facilitator notes

For each slide write SAY, ASK, DO, WATCH FOR, DEBRIEF, expected responses, transition, time, optional cut, virtual adaptation, and fallback. Preserve approved sources verbatim. Apply Shared safeguards.

10. Assessment creation

For each objective create baseline item, formative check, post-performance task, transfer measure, answer key/rubric, and construct-alignment explanation. Avoid trivia and satisfaction proxies. Apply Shared safeguards.

11. Misconception identification

List likely misconceptions, why each is attractive, correcting evidence, diagnostic question, facilitator response, and prevention activity. Do not invent author beliefs. Apply Shared safeguards.

12. Accessibility review

Audit cognitive load, plain language, reading order, contrast, color-only meaning, alt text, captions, keyboard use, motion, screen-reader structure, bandwidth, language fluency, AI familiarity, and cultural assumptions. Return issue, impact, exact fix, owner, priority.

13. Content QA

Compare all assets. Flag contradictory numbers, terminology drift, missing sources, alignment gaps, exposed answers, timing conflicts, outdated versions, unsupported promises, and disclaimer inconsistencies. Produce a release-blocking issue list; do not silently fix.

14. Feedback analysis

Analyze de-identified feedback, rubric scores, timing logs, and facilitator notes. Preserve dissent and accessibility signals. Return themes, frequency, severity, evidence, performance correlation, keep/change/test decision, confidence, and data needed next. Do not equate preference with effectiveness.

15. Curriculum revision

Propose the smallest evidence-based revisions. For each include problem, evidence, asset/location, exact change, introduced risk, approval, version increment, and next-cohort validation. Do not alter research interpretation without faculty approval.

Release note template

Project ID: ____ Release: ____ Date: ____ Owner: ____

Research source version: ____ Faculty approval: ____

Assets: ____

Changes since prior version: ____

Evidence motivating change: ____

Known limitations: ____

Open validation items: ____

Accessibility status: ____

Delivery dependencies/fallbacks: ____

Approval signatures: ____

Facilitator observation log

Actual segment times: ____

Hinge-question result: ____

Common misconception: ____

Where participants stalled: ____

Percentage with task-level map: ____

Percentage with representative edge case: ____

Percentage with true stop rule: ____

Accessibility/participation issue: ____

Suggested change and evidence: ____

Requires faculty review? ____

Closing script

“The frontier is not a line someone else can map once for your organization. It moves by task, context, model, and consequence. Your advantage is not predicting it perfectly. It is building a disciplined way to test, assign judgment, and learn faster than the risk travels. In the next seven days, I will test ___ before allowing AI to ___.”

Sources

- Harvard Careers - Instructional Designer, HBS AI Institute
- Navigating the Jagged Technological Frontier - peer-reviewed paper
- HBS AI Institute - Back to the Beginnings of AI at Work
- HBS AI Institute - Accelerate with AI
- HBS AI Institute - Future Proof with AI