From HBS Research to Experiential AI Learning

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

Executive summary

This proof-of-work tests a specific question: can rigorous AI research become a short executive-learning experience that changes how leaders make decisions at work?

The prototype starts with the live Instructional Designer, HBS AI Institute job description, evaluates nine public HBS or HBS-affiliated AI research candidates, and selects the peer-reviewed study *Navigating the Jagged Technological Frontier*. It then translates the research into a 60-minute workshop, a reusable decision framework, an organization-specific exercise, a live demonstration, assessments, a feedback loop, an AI-enabled production workflow, an SOP, a prompt library, multimedia specifications, an optional digital companion, a portfolio case study, and interview materials.

The design principle is simple: participants should not leave merely knowing that AI is uneven. They should leave able to decompose a workflow, test where AI helps or harms, assign an appropriate human-AI operating mode, and launch a measured seven-day experiment.

1. Current HBS job-description analysis

Source of record: Harvard Careers - Instructional Designer, HBS AI Institute, posting 002202SR, posted May 13, 2026.

The posting describes the Institute as helping industry leaders reimagine AI in business and society through research and experiential learning. It positions the designer as the connector among faculty, subject-matter experts, facilitators, and go-to-market teams, responsible for the end-to-end learning-asset lifecycle. The Institute's public enterprise-learning site reinforces a learning-by-doing model built around case discussion and workshop activities, while its digital platform combines short modules, guided exercises, faculty perspectives, and practical application. Sources: Accelerate with AI, Future Proof with AI, and HBS AI Institute overview.

HBS requirement	What it means in practice	Artifact that proves capability
Connect faculty, SMEs, facilitators, and go-to-market teams	Translate accurately across academic, learning, delivery, and market contexts; make approvals and handoffs visible	Faculty research brief; RACI and approval gates in the SOP; microsite copy with an explicit prototype disclaimer
Manage the end-to-end lifecycle of experiential learning assets	Own intake through validation, build, delivery readiness, measurement, revision, and archive	AI-assisted production workflow; production SOP; release checklist; version-history model
Rapidly convert program materials into client-ready and market-facing assets with AI	Use AI to accelerate transformation while preserving evidence, voice, and intent	Deck, workbook, guide, video script, infographic specification, takeaway card, microsite copy, and prompt library
Ensure quality, coherence, and consistency through version control, QA, and release documentation	Establish a single source of truth, trace claims, test files, and document what changed	Source ledger; QA gates; release manifest; change log; final quality audit
Partner with faculty and SMEs to refresh facilitation materials, exercises, and learning assets	Treat faculty interpretation as authoritative and build an efficient review loop	Faculty brief; validation questions; SME review checkpoints; “requires faculty validation” flags
Translate complex research into accessible, participant-centered formats	Preserve findings and limitations while converting implications into learner decisions and actions	Research-to-learning rationale; Frontier Fit Loop; workshop deck; workbook; experiential exercise
Create decks, toolkits, videos, blog posts, infographics, and digital content	Select formats by learning need rather than novelty	Facilitation deck; executive card; multimedia scripts/specifications; case-study and microsite copy
Support delivery through preparation, facilitator support, asset readiness, and post-session refinement	Make delivery repeatable across facilitators and resilient to technology failure	Minute-by-minute guide; SAY/ASK/DO/WATCH FOR prompts; contingency plans; asset-readiness checklist
Incorporate participant and facilitator feedback	Combine perception, performance, observation, and transfer data into prioritized revisions	Feedback instrument; cohort analysis example; revision decision rules
Develop metrics for learning performance, engagement, and impact	Measure ability and transfer, not only attendance or satisfaction	Before/during/after/later assessments; application rubric; 7/30/90-day measures
Develop SOPs for AI-enabled digital-learning production and delivery	Turn one-off craft into an auditable, repeatable operating system	SOP: AI-Assisted Experiential Learning Production
Manage complex virtual programming and support in-person delivery	Design for room, remote, hybrid, and low-tech conditions	Setup plans; technology checklist; virtual participation pattern; offline fallback
Translate technical concepts for non-technical audiences	Use plain language and concrete decisions without distorting the underlying technology	Faculty brief; executive mental model; live demo; glossary and misconceptions
Choose, configure, and apply AI tools responsibly	Match tools to tasks, control source context, protect data, and require accountable review	Human/AI responsibility matrix; evidence controls; prompt safeguards; privacy rules
Use content, video, workflow, automation, and agent tools	Build a production system rather than isolated prompts	Prompt library; automation map; digital-companion system prompt; revision pipeline
Demonstrate working knowledge of AI/ML concepts and architectures	Explain model limits, data dependence, evaluation, retrieval, fine-tuning, and agentic workflows at an appropriate level	AI-production notes; digital-companion architecture; facilitator difficult-questions section
Build trust through on-site collaboration	Make uncertainty visible, invite correction, and keep humans accountable	Collaborative walkthrough language; faculty-validation list; decision log and approval trail

Design specification derived from the role

The project is successful only if it demonstrates five connected capabilities:

1. Research fidelity - claims trace to primary or official HBS sources; findings, implications, and prototype assumptions remain distinct.
2. Experiential design - the session alternates Learn, See, Do, Reflect, and Apply, and participants work on a real organizational problem.
3. Delivery readiness - a facilitator can run the session without the designer present, including when the live AI fails.
4. Measurement - performance and transfer are observable through aligned assessments.
5. Scalable production - AI accelerates transformation and analysis inside explicit human and faculty approval gates.

2. HBS AI research landscape

Scoring scale: 1 = weak fit; 10 = exceptional fit. Scores are prototype judgments, not Harvard ratings.

Research candidate and source	Core finding and executive value	Potential learning objective	Experiential exercise concept
Navigating the Jagged Technological Frontier - Dell'Acqua, McFowland, Mollick, Lifshitz, Kellogg, Rajendran, Krayer, Candelon, Lakhani. Peer-reviewed paper	AI improved speed, completion, and quality on tasks inside its capability frontier but reduced correctness on a task outside it. Leaders need workflow-level tests and guardrails, not blanket adoption rules.	Design a representative task-level comparison and assign AI/human authority from evidence.	Map a real workflow, run a representative test, assign roles, and define a stop rule.
The Cybernetic Teammate - Dell'Acqua and colleagues including Sadun and Lakhani. HBS AI Institute summary	In the studied P&G product-development task, AI-enabled individuals could reach team-level quality, reduce time, and bridge functional silos.	Select a team configuration appropriate to task and expertise integration.	Redesign a decision as solo, human team, AI-enabled individual, or AI-enabled team.
Mapping AI into Production - Kim, Kim, Koning. HBS AI Institute summary	In a field experiment with 515 startups, organizational examples broadened the search for AI uses and produced differences in use cases and firm outcomes. Access is not the same as knowing where to apply AI.	Locate a production bottleneck and identify complementary workflow changes.	Map the full production chain, locate the binding constraint, and redesign connected tasks.
The Uneven Impact of Generative AI on Entrepreneurial Performance - Otis, Clarke, Delecourt, Holtz, Koning. HBS AI Institute summary	Average effects hid divergence: higher baseline performers benefited while lower performers were harmed in the studied setting, apparently through selection and implementation differences.	Evaluate advice quality, contextual fit, and implementation evidence.	Compare the same AI advice across two contexts; design a selection-and-implementation checkpoint.
Narrative AI and the Human-AI Oversight Paradox - Lane and collaborators. HBS AI Institute summary	AI recommendations changed innovation-screening stringency; narrative rationales increased influence, sharpening oversight needs for subjective criteria.	Calibrate human oversight to objective and subjective criteria.	Evaluate proposals with no AI, recommendation-only AI, and narrative AI.

Research candidate and source	Core finding and executive value	Potential learning objective	Experiential exercise concept
Cyborgs, Centaurs and Self-Automators - Randazzo and colleagues. HBS AI Institute summary	More than 200 consultants clustered into fluid co-creation, directed co-creation, and delegated/abdicated modes. Operating pattern matters as much as access.	Diagnose and deliberately choose a collaboration mode.	Redesign one task for a different human-AI mode.
Innovating with Generative AI: A Human Bottleneck Framework - De Freitas, Israeli, Nave, Timoshenko, Toubia. HBS paper	AI may alleviate, shift, or amplify human bottlenecks across innovation stages. Faster ideation does not remove failures in noticing, evaluating, choosing, or acting.	Diagnose whether an AI intervention removes or relocates a bottleneck.	Map one innovation stage and test the next bottleneck created.
Novice Risk Work - Kellogg, Lifshitz-Assaf, Randazzo, Mollick, Dell'Acqua, McFowland, Candelon, Lakhani. HBS faculty record	Junior employees acting as de facto coaches on emerging technology can produce learning failures when technical understanding is shallow.	Design enablement roles and escalation paths that protect learning quality.	Map who experiments, validates, teaches, approves, and escalates.
The Innovation Race - team including HBS AI Institute associate Zoe Cullen. HBS AI Institute summary	A field experiment with roughly 3,000 Italian firms found that correcting beliefs about peer adoption affected plans differently for robotics and AI.	Separate evidence, competitor belief, and strategic fit in an adoption decision.	Audit an AI investment decision for imitation versus operating logic.

Candidate scoring

Candidate	RQ	ER	EL	PU	JD	ID
Jagged Technological Frontier	10	10	10	10	10	10
Cybernetic Teammate	10	10	9	9	10	9
Mapping AI into Production	9	10	10	10	10	10
Uneven Entrepreneurial Performance	9	10	9	10	10	9
Narrative AI / Oversight Paradox	9	9	10	9	10	10
Cyborgs, Centaurs, Self-Automators	8	10	10	9	9	10
Human Bottleneck Framework	8	10	9	9	9	9
Novice Risk Work	9	9	9	9	9	9
Innovation Race	9	9	8	8	8	8

Abbreviations: RQ = research quality; ER = executive relevance; EL = experiential-learning potential; PU = practical usefulness; JD = job-description alignment; ID = ability to demonstrate instructional-design capability.

3. Selected research and justification

Selection

Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality, formally published in *Organization Science* in 2026.

Primary source: peer-reviewed paper. Accessible official interpretation: HBS AI Institute, "Back to the Beginnings of AI at Work".

Why it wins

- It is peer reviewed, preregistered, randomized, and conducted with 758 working consultants on tasks designed to resemble their work.
- It creates a decision problem, not a lecture topic: leaders must determine where AI supports or harms a workflow even when the boundary is hard to see.
- The findings support a high-value real-work exercise: decomposition, testing, role assignment, guardrails, and measurement.
- The concept is durable. Specific models will change, but an uneven and moving capability frontier still makes workflow-level evaluation necessary.
- It creates space to demonstrate research fidelity. The workshop must resist two seductive overclaims: that AI always boosts productivity and that every task can be safely classified in advance.
- It maps directly to the role: faculty translation, facilitation assets, measurement, digital continuation, AI-assisted production, and continuous improvement.

Decision not to select the newest research

The June 2026 Human Bottleneck Framework and March 2026 Mapping AI into Production study are highly relevant. The first is a conceptual working paper and the second is a new working paper. Jagged Frontier has the advantage of peer-reviewed publication, a causal design, memorable countervailing findings, and a cleaner 60-minute experiential arc. A future learning series could sequence the three: locate the frontier, map AI into production, then diagnose the human bottlenecks that remain.

4. Faculty Research Brief

Research question

How does access to a frontier generative-AI system affect the productivity and quality of highly skilled knowledge workers, and how do results change when a task is inside versus outside the system's capability frontier?

What was actually investigated

The researchers collaborated with Boston Consulting Group on a preregistered randomized laboratory-in-the-field experiment involving 758 consultants. Participants completed a baseline assessment and were assigned to

one of three conditions: no AI, GPT-4 access, or GPT-4 access plus a brief prompt-engineering overview. They completed two sets of realistic consulting-style work: an 18-subtask product innovation and go-to-market assignment designed to be within GPT-4's frontier, and a brand-strategy case designed to require reconciliation of quantitative and qualitative evidence in a way the model did not reliably handle. The model was the April 2023 version of GPT-4, used in June 2023 through a custom interface. Primary paper.

Evidence

- Randomized assignment and preregistration support causal interpretation within the experimental setting.
- Tasks were developed with the partner organization to resemble consultant workflows.
- Inside-frontier quality was evaluated by independent human graders; the study also reported robustness checks using other evaluation methods.
- Outside-frontier correctness was evaluated against the intended solution and checked by BCG consultants.
- Logged human-AI interactions made it possible to examine how participants used the tool.

Findings

Inside the frontier, AI users completed 12.2% more subtasks and worked 25.1% faster on average; the HBS AI Institute's 2026 interpretation reports work rated about 32% higher in quality. Both lower- and higher-baseline performers benefited in the studied task, with larger gains for the lower half. Outside the frontier, the no-AI group was correct 84.5% of the time, compared with 70.6% for GPT-only and 60% for GPT plus the overview; combined, AI users were about 19 percentage points less likely to reach the correct answer. Their outputs could still appear coherent and persuasive. HBS AI Institute summary and primary paper.

What the findings mean

AI's effect is task dependent. Workflows can cross the capability boundary several times, and that boundary may not match human intuitions about task difficulty. Leaders therefore need task-level evaluation and evidence-aware human oversight.

Limitations

- The experiment tested one version of GPT-4 available in 2023. The location of the frontier moves as models, tools, context windows, retrieval, and workflows change.
- Participants were highly selected consultants; results should not be assumed to transfer unchanged to every occupation, geography, organization, or skill level.
- The tasks resembled real work but did not reproduce every consequence, dependency, time horizon, or social complexity of live client work.
- The study deliberately pretested tasks to sit on different sides of the frontier. In practice, leaders usually do not know the classification beforehand.
- The research does not supply a universal list of tasks AI should automate, nor does it establish that prompt training is harmful in general. The lower outside-frontier accuracy in the prompt-overview condition occurred in this design and should not be generalized without further evidence.
- Productivity and quality gains in a bounded assignment do not by themselves establish long-term organizational value, learning retention, workforce effects, or risk.

Executive implications - prototype interpretation

1. Replace blanket adoption decisions with workflow decomposition and task-level tests.
2. Treat polished, coherent output as a risk signal when correctness is hard to observe.
3. Preserve independent human judgment for consequential, ambiguous, novel, or hard-to-verify work.
4. Measure quality and decision correctness alongside speed and volume.
5. Re-test as models, data, policies, and workflows change.

These are instructional translations, not claims that the authors prescribed this exact operating model. Requires HBS/faculty validation.

Likely misinterpretations

- Incorrect: “AI improves knowledge work by roughly 30%.” Correct: the observed effect depended on task position relative to the tested model's frontier and on the outcome measured.
- Incorrect: “Lower performers always gain the most.” Correct: that pattern appeared inside the frontier in this sample; other HBS-affiliated research has found different heterogeneous effects in open-ended entrepreneurial settings.
- Incorrect: “Humans can simply check AI.” Correct: checking is difficult when the output is persuasive, the error is not obvious, or the human has already anchored on the AI's framing.
- Incorrect: “The answer is better prompting.” Correct: the study concerns task fit and human-AI workflow design; prompting alone did not remove outside-frontier failure.
- Incorrect: “The frontier is fixed.” Correct: it moves with models and configurations, which makes monitoring and revalidation necessary.

Transferable principles

- Unit of analysis: the task within a workflow, not the job title or tool.
- Evidence before scale: compare AI-assisted performance with a credible baseline on representative cases.
- Confidence is not correctness: fluent output needs independent evidence.
- Match control to consequence: as reversibility falls and stakes rise, human authority and verification should increase.
- Revalidation is operational maintenance: a model, data, or process change can move the frontier.

5. Learner personas

Persona A - Functional executive sponsor

Role: CMO, COO, CHRO, CFO, or business-unit leader. AI fluency: low to moderate. Motivation: credible productivity and competitive advantage. Fear: reputational harm, uncontrolled tools, wasted transformation budget. Misconception: AI strategy means selecting a platform and setting a use policy. Business problem: many pilots, weak evidence of value. Needs: a way to ask sharper questions of teams. Value signal: leaves with a decision rule and one well-scoped experiment. Disengagement trigger: product tutorials, jargon, or exercises detached from strategic decisions.

Persona B - Operating manager and workflow owner

Role: director or senior manager responsible for a repeatable process. AI fluency: moderate and uneven. Motivation: reduce cycle time without increasing defects. Fear: becoming accountable for invisible AI errors. Misconception: if a demonstration works once, the workflow is ready. Business problem: inconsistent inputs, handoffs, exceptions, and approval delays. Needs: workflow decomposition, test cases, guardrails, and metrics. Value signal: converts a real workflow into a seven-day test plan. Disengagement trigger: abstract ethics debate with no operating choices.

Persona C - Founder or general manager

Role: founder, GM, or small-business operator. AI fluency: moderate to high through self-experimentation. Motivation: increase capacity without proportional headcount. Fear: moving too slowly or building brittle automation. Misconception: speed of output equals leverage. Business problem: too many opportunities and too little evaluation discipline. Needs: prioritization and stop rules. Value signal: identifies where not to automate. Disengagement trigger: lengthy governance language before a practical tool appears.

Persona D - Risk, people, or enablement leader

Role: legal, compliance, HR, learning, or transformation leader. AI fluency: low to moderate. Motivation: responsible adoption and capability building. Fear: sensitive-data leakage, bias, deskilling, and ambiguous accountability. Misconception: guardrails are a policy document rather than a workflow behavior. Business problem: translating broad principles into everyday decisions. Needs: a graduated control model and observable evidence. Value signal: sees how to pair learner autonomy with escalation and review. Disengagement trigger: false certainty or a “move fast” culture that dismisses risk.

Inclusive design implication

The session must allow the same framework to work for high-growth experimentation and regulated decision-making. Participants choose their own workflow, use plain-language definitions, and can complete the core exercise without live AI access.

6. Learning objectives through backward design

After the experience, participants can…	Knowledge required	Skill required	Development activity	Assessment evidence
Decompose one business workflow into observable tasks and decision points	Workflow versus task; inputs, outputs, handoffs, exceptions	Define task boundaries at a level that can be tested	Workflow strip and task decomposition	At least 4 distinct tasks with owners, inputs, outputs, and failure consequences
Form an evidence-based hypothesis about where AI may help, harm, or remain uncertain	Jagged frontier; coherence versus correctness; moving boundary	Identify observability, repeatability, context dependence, and stakes	Frontier prediction on each task	Rationale cites task characteristics rather than personal enthusiasm
Design a representative comparison test for one task	Baseline, representative case, acceptance criteria, failure mode	Compare human-only and human-AI outputs using the same rubric	Test-plan canvas and demo audit	Test includes baseline, 3-5 cases, quality metric, and stop rule

After the experience, participants can...	Knowledge required	Skill required	Development activity	Assessment evidence
Assign an appropriate AI-human operating mode and guardrail	AI-led, human-AI, human-led; verification and escalation	Match authority to evidence, uncertainty, and consequence	Mode decision matrix and peer challenge	Role assignment includes accountable human, verification source, and escalation condition
Commit to a seven-day experiment that can produce a go, revise, or stop decision	Transfer, iteration, monitoring	Scope a small, reversible experiment with meaningful measures	Action-plan card	Named owner, first case, metric, review date, and decision rule

7. Learning architecture

Communication job: by the end, executives and managers should be able to test one real workflow and assign AI an evidence-based role because the effect of AI is jagged, moving, and task dependent.

Time	Mode	Learner experience	Output
0-5	Learn / predict	Provocation: “Would you deploy a tool that is faster and more persuasive but sometimes less correct?” Individual vote and pair rationale	Prior belief made visible
5-12	Learn	Research story: design, inside-frontier gains, outside-frontier losses, and limitations	Accurate mental model
12-20	See / reflect	Frontier Fit Loop introduced and applied to a short familiar example	Framework comprehension
20-30	See / do	Live “polished but wrong” decision demo; participants identify the verification failure	Concrete experience of overreliance
30-45	Do	Participants map a real workflow, choose one task, and build a representative test	Frontier Fit Map and test plan
45-53	Reflect	Peer challenge and facilitated debrief; compare assumptions, evidence, and risk	Revised decision
53-58	Apply	Build a seven-day experiment with owner, metric, stop rule, and review	Transfer plan
58-60	Commit	One-sentence commitment and post-assessment	Public next action

Alternation: Learn -> Predict -> Learn -> See -> Do -> Reflect -> Apply -> Commit. No passive segment exceeds seven minutes.

8. Signature framework - the Frontier Fit Loop

The framework

Decompose -> Predict -> Test -> Assign -> Monitor

1. Decompose the workflow into tasks with observable inputs, outputs, decisions, and failure consequences.

2. Predict frontier fit: promising, uncertain, or restricted. Record the reason before seeing the AI output to reduce hindsight bias.
3. Test representative cases against a human-only or current-process baseline. Score correctness, usefulness, time, and failure severity.
4. Assign the operating mode: AI-led with review, human-AI collaboration, or human-led with optional AI support. Name the accountable human and evidence source.
5. Monitor drift, exceptions, overrides, and business impact. Re-test after meaningful changes to the model, data, prompt, policy, or process.

Derivation from the research

- Decompose follows the paper's finding that one workflow can contain tasks on both sides of the frontier.
- Predict acknowledges that task position is difficult to know in advance and makes assumptions inspectable.
- Test responds to the observed performance reversal and prevents coherent output from serving as proof.
- Assign operationalizes the paper's emphasis on human judgment without pretending every task needs the same amount of oversight.
- Monitor reflects the paper's description of an expanding, changing frontier.

The labels and operating procedure are original to this candidate prototype. They are not a framework named or endorsed by the paper's authors. Requires HBS/faculty validation.

Operating-mode decision rule

Evidence and consequence	Recommended prototype mode
Repeated strong performance; errors easy to detect and reverse; low consequence	AI-led execution with sampled human review
Mixed or incomplete evidence; important context; moderate consequence	Human-AI collaboration with explicit independent verification
Weak evidence; errors hard to detect; novel, high-stakes, regulated, or irreversible decision	Human-led decision; AI may support bounded analysis but cannot hold final authority

9. Complete workshop deck - slide-by-slide specification

Slide 1 - Where does AI help - and where does it quietly hurt?

Purpose: open with the decision problem. On-screen content: title; “A 60-minute executive working session”; prototype disclaimer. Suggested visual: a single uneven boundary line. Notes: establish that the session is about workflow decisions, not model features. Participant action: choose a real workflow to use later.

Slide 2 - A faster answer can still be the wrong answer

Purpose: provoke judgment. On-screen content: “25% faster. More persuasive. Less correct. Deploy?” Suggested visual: three large figures with the last in a caution color. Notes: figures come from different task conditions in

the study; do not imply the same task produced all three at once. Participant action: vote yes/no/not enough evidence and give one reason.

Slide 3 - Your outcome today

Purpose: make performance expectation explicit. On-screen content: “Leave with a tested hypothesis for one task, one human-AI operating mode, and one seven-day experiment.” Suggested visual: three verbs - Test / Assign / Act. Participant action: write the workflow they brought.

Slide 4 - The study asked a practical question

Purpose: introduce research question. On-screen content: “What happens when highly skilled knowledge workers use AI on realistic work?” Suggested visual: human, task, AI, outcome. Notes: preregistered randomized laboratory-in-the-field experiment with 758 BCG consultants. Participant action: listen.

Slide 5 - Same professionals. Different task fit.

Purpose: explain conditions. On-screen content: “No AI / GPT-4 / GPT-4 + brief overview”; “Innovation and go-to-market work”; “Brand-strategy diagnosis.” Suggested visual: two task lanes. Notes: clarify that tasks were pretested to be on different sides of the then-current frontier. Participant action: predict which task produced the risk.

Slide 6 - Inside the frontier, AI raised performance

Purpose: present positive evidence. On-screen content: “+12.2% subtasks completed”; “25.1% faster”; “about 32% higher quality.” Suggested visual: three large figures with a short “in this task and study” qualifier. Notes: quality figure follows the HBS AI Institute 2026 summary. Participant action: identify which business metric each result resembles.

Slide 7 - Outside the frontier, accuracy fell

Purpose: present counterevidence. On-screen content: “No AI 84.5% correct”; “GPT-4 70.6%”; “GPT-4 + overview 60%.” Suggested visual: descending bars. Notes: emphasize that the AI handled surface-level information but missed a critical qualitative clue in the designed case. Participant action: note what makes this failure dangerous.

Slide 8 - Coherent is not the same as correct

Purpose: surface the core risk. On-screen content: “Fluency can hide a verification failure.” Suggested visual: polished memo with an evidence gap highlighted. Notes: the paper reports outside-frontier recommendations could be rated as coherent even when the conclusion was wrong. Participant action: share one domain where a polished error travels far.

Slide 9 - The frontier is jagged

Purpose: teach the mental model. On-screen content: “Similar-looking tasks can sit on opposite sides of AI capability.” Suggested visual: uneven boundary crossing a workflow. Notes: avoid a fixed capability map; the boundary changes. Participant action: mark one task they think is promising and one uncertain.

Slide 10 - What this research does not prove

Purpose: preserve academic integrity. On-screen content: “Not every model. Not every job. Not a permanent boundary. Not permission to automate by intuition.” Suggested visual: four short limits. Notes: mention the 2023 GPT-4 version, selected consultant sample, and simulated work. Participant action: ask one clarification.

Slide 11 - The Frontier Fit Loop

Purpose: introduce transfer tool. On-screen content: “Decompose -> Predict -> Test -> Assign -> Monitor.” Suggested visual: circular or horizontal five-step loop. Notes: label as candidate-created translation requiring faculty validation. Participant action: read steps aloud in pairs.

Slide 12 - Decompose the workflow, not the job

Purpose: model useful granularity. On-screen content: “Input -> task -> decision -> output -> handoff”; example: “Customer escalation: summarize -> diagnose -> choose remedy -> write response -> approve.” Suggested visual: simple workflow strip. Participant action: split their workflow into 4-7 tasks.

Slide 13 - Predict before the demo changes your mind

Purpose: reduce hindsight bias. On-screen content: “Promising / uncertain / restricted”; cues: observable, repeatable, contextual, consequential. Suggested visual: three labels on a continuum. Notes: these are hypotheses, not final classifications. Participant action: classify one task and record why.

Slide 14 - Live demo: choose a launch market

Purpose: create a concrete decision. On-screen content: two fictional markets; structured metrics favor Northstar; interview notes reveal a non-transferable distribution agreement. Suggested visual: small evidence split. Notes: use only fictional, non-confidential data. Participant action: independently choose a market before AI is shown.

Slide 15 - The AI answer looks excellent

Purpose: expose persuasive failure. On-screen content: a concise AI recommendation for Northstar based on market size and acquisition cost. Suggested visual: clean recommendation excerpt. Notes: use a prepared output if live AI differs or fails. Participant action: rate confidence 1-5.

Slide 16 - The missing evidence changes the decision

Purpose: demonstrate independent verification. On-screen content: “Constraint: distribution agreement cannot transfer”; “Revised decision: run a bounded feasibility test before launch.” Suggested visual: red evidence flag connected to the decision. Participant action: identify which Frontier Fit step failed.

Slide 17 - Your Frontier Fit Map

Purpose: launch the main exercise. On-screen content: “Choose one consequential, repeatable task. Map it. Test it. Assign a mode.” Suggested visual: workbook page thumbnail. Notes: provide a low-risk sample for participants without a workflow. Participant action: turn to worksheet.

Slide 18 - Step 1: map the current task

Purpose: guide work. On-screen content: “Desired outcome / inputs / repeated decisions / human judgment / failure consequence / current baseline.” Suggested visual: six prompts. Participant action: complete the first canvas.

Slide 19 - Step 2: design a fair test

Purpose: prevent demo theater. On-screen content: “Same cases. Same rubric. Human/current baseline. 3-5 representative examples. Stop rule.” Suggested visual: A/B comparison. Participant action: write test cases and acceptance criteria.

Slide 20 - Step 3: assign authority, not just tasks

Purpose: convert evidence into operating design. On-screen content: “AI-led / Human-AI / Human-led”; “Who decides? What evidence? When escalate?” Suggested visual: three operating modes. Participant action: select a mode and name accountable human.

Slide 21 - Peer challenge

Purpose: improve decision quality. On-screen content: “Where could a polished error survive? Is the test representative? Is the metric meaningful? Is the action reversible?” Suggested visual: four questions. Participant action: exchange maps and challenge one assumption.

Slide 22 - Debrief: the pattern is the lesson

Purpose: connect activity to research. On-screen content: “Where did you move the human checkpoint - and why?” Suggested visual: before/after checkpoint. Notes: compare tasks, not industries; emphasize uncertainty and evidence. Participant action: share one design change.

Slide 23 - Your seven-day experiment

Purpose: enable transfer. On-screen content: “Owner / first case / baseline / success metric / stop rule / review date.” Suggested visual: one-page experiment card. Participant action: complete and calendar the review.

Slide 24 - One commitment

Purpose: close with action. On-screen content: “In the next seven days, I will test ___ before allowing AI to ___.” Suggested visual: blank line. Notes: collect post-assessment; remind participants that re-testing is part of responsible use. Participant action: state or submit commitment.

10. Participant workbook

The companion workbook is designed as a durable operating tool, not slide printouts. It includes:

1. Workshop outcome and research-source disclaimer.
2. One-page research brief with findings and limitations.
3. Frontier Fit Loop reference page.
4. Baseline scenario and confidence check.
5. Workflow selection prompt.
6. Workflow decomposition strip.
7. Frontier prediction worksheet.
8. Representative test-plan canvas.
9. Operating-mode decision tool.
10. Human checkpoint and escalation design.
11. Seven-day experiment card.
12. Completed example.
13. Reflection and personal action plan.
14. One-page executive takeaway card.

Actual worksheet copy appears in the separate participant workbook artifact.

11. Main experiential exercise - Frontier Fit Map

Participant instructions

Choose a recurring workflow that matters, has at least four steps, and could be tested without exposing confidential or regulated data. Avoid using a life-or-death or legally dispositive decision for the first experiment.

1. Name the desired business outcome and current baseline.
2. Break the workflow into 4-7 tasks. For each, record inputs, output, decision, owner, and failure consequence.
3. Choose one task. Predict its frontier fit: promising, uncertain, or restricted. Explain why.
4. Create 3-5 representative test cases, including at least one edge case.
5. Define a comparison: current process or human-only versus human-AI.

6. Define acceptance criteria for correctness/usefulness, time, and failure severity.

7. Choose an operating mode and human checkpoint.

8. Write a stop rule and a seven-day next action.

Participant worksheet

Workflow: ____ Desired outcome: ____ Current cycle time/quality: ____

Task 1: ____ Input: ____ Decision: ____ Output: ____ Owner: ____ Failure consequence: ____

Repeat for Tasks 2-7.

Selected task: ____ Prediction: Promising / Uncertain / Restricted

Reason: Observable? ____ Repeatable? ____ Context-dependent? ____ Consequential? ____

Test cases: Typical 1 ____ Typical 2 ____ Edge case ____ Adversarial/ambiguous case ____

Baseline: ____ AI-assisted condition: ____ Blind scoring method: ____

Metrics: correctness/quality ____ time ____ error severity ____ override rate ____

Operating mode: AI-led / Human-AI / Human-led

Accountable human: ____ Verification source: ____ Escalate when: ____ Stop when: ____

Seven-day action: ____ Owner: ____ Review date: ____ Go/revise/stop rule: ____

Facilitator instructions

Keep participants at task level. If someone writes “marketing” or “finance,” ask, “What observable action begins and ends?” If someone selects a high-stakes task, keep it as a design exercise and require a safer test environment. At minute 37, prompt participants to add an edge case. At minute 42, require a named evidence source. Pair participants across functions when possible.

Common mistakes

- Mapping a job or department rather than a task.
- Treating frequent use as evidence of quality.
- Testing only the easiest cases.
- Measuring speed but not correctness or failure severity.
- Naming “human review” without saying what the human checks or what evidence they use.
- Giving AI final authority because the output is well written.
- Treating the first classification as permanent.

Discussion prompts

- Which error could survive your current review process?
- What would make the AI look successful while the business outcome worsens?
- What data or context exists only in people's heads?
- When does the cost of verification erase the speed gain?

- What would have to be true before you loosen the human checkpoint?

Debrief

Return to the research: the same professionals using the same model experienced different performance effects on different tasks. The exercise does not locate a permanent frontier. It creates a disciplined way to learn where the frontier appears to be for a specific task, configuration, and consequence.

Extension

Run the map across an entire value chain. Compare local task gains with end-to-end throughput. Identify the slowest or riskiest complementary step, then redesign the workflow rather than optimizing isolated tasks.

12. Completed exercise example

Scenario

A mid-sized B2B software company wants to reduce the time managers spend triaging customer escalations.

Current workflow

Task	Input	Decision/output	Human judgment	Failure consequence
Gather case history	Tickets, account notes, product logs	Consolidated record	Resolve conflicting records	Missing context
Summarize issue	Consolidated record	One-page summary	Distinguish symptoms from cause	Misframed problem
Classify severity	Summary, SLA, account tier	P1-P4 classification	Interpret business impact and exceptions	Wrong response level
Recommend remedy	Classification, playbook	Action options	Balance customer promise, product risk, precedent	Cost, churn, or security exposure
Draft response	Approved remedy	Customer message	Tone, commitments, relationship context	Reputational or contractual risk
Approve and send	Full packet	Final decision	Accountability	Irreversible external commitment

Selected task and prediction

Selected task: summarize issue. Prediction: promising, because the output is observable and repeatable, but uncertain where records conflict or causality is ambiguous.

Test

Ten de-identified historical cases: six typical, two with conflicting notes, one security-sensitive, one with missing logs. Compare current manager summary with AI-assisted summary. Two blinded reviewers score factual

completeness, unsupported claims, clarity, and correct identification of uncertainty. Record time. Stop if the AI invents a customer commitment, exposes restricted data, or omits a security flag.

Operating mode

Human-AI collaboration. AI drafts a structured summary; the case owner verifies each fact against the source packet and marks unresolved conflicts. AI cannot classify severity or recommend an external commitment in this pilot.

Metrics and decision rule

Proceed to a wider pilot only if median cycle time improves by at least 20%, factual-completeness score is no worse than baseline, unsupported-claim rate is zero on critical facts, and reviewers identify every security-sensitive case. Otherwise revise or stop.

13. Live demonstration

Scenario

“Northstar or Harbor?” A fictional consumer-subscription company must choose a market for a six-week launch test.

Starting problem and inputs

Structured metrics favor Northstar: larger reachable audience, lower acquisition cost, and faster projected payback. Interview notes contain a critical constraint: the company's current distribution agreement is valid only in Harbor, and Northstar approval would take 10-14 weeks. The executive decision is not simply which market is more attractive; it is which experiment can actually launch within six weeks.

Prompt 1 - intentionally insufficient

“Act as a growth strategist. Review the market metrics below and recommend Northstar or Harbor for a six-week launch test. Provide a decisive recommendation, three reasons, key risks, and a first-week action plan. Metrics: Northstar reachable audience 220,000, CAC $42, projected 9-month contribution $1.4M. Harbor reachable audience 150,000, CAC $58, projected contribution $0.9M.”

Expected prepared output

The AI recommends Northstar because of larger audience, lower acquisition cost, and higher projected contribution. The answer is clear, plausible, and incomplete because the input excluded the decisive operational constraint.

Prompt 2 - full evidence but weak verification

“Update the recommendation using these interview notes: Legal says our current distribution agreement covers Harbor. Northstar approval normally requires 10-14 weeks. Sales says a channel partner might accelerate approval but there is no written commitment. Keep the six-week launch objective.”

Expected output

A well-performing model should revise toward Harbor or a bounded feasibility test. If it does not, the demonstration becomes even more useful: ask participants to identify the unsupported assumption. Because model output is nondeterministic, the facilitator keeps a prepared Northstar recommendation and a prepared corrected analysis.

Where AI succeeds

It organizes comparable metrics, articulates tradeoffs, drafts options, and updates quickly when context is added.

Where it fails

It cannot use evidence it was not given; it may overvalue clean quantitative data, accept an unwritten possibility as a plan, or produce unwarranted certainty.

Human judgment required

Define the actual decision, determine which evidence is authoritative, verify contractual facts, distinguish option value from commitment, and decide whether a six-week test or a longer market entry is the real objective.

Discussion question

“Which Frontier Fit step would have prevented the first polished answer from becoming a business decision?”

14. Facilitator guide summary

The separate guide contains preparation, room and virtual setup, complete timing, SAY/ASK/DO/WATCH FOR/DEBRIEF language, expected responses, misconceptions, difficult questions, accessibility accommodations, timing variants, and technology contingencies.

Core facilitator stance

- Curious rather than prescriptive: “What evidence would change your classification?”
- Precise about research: “In this experiment and task,” not “AI always.”
- Protective of learner context: no confidential data in public models.
- Focused on decisions: bring tool discussion back to task, evidence, authority, and outcome.

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. Complete your sentence: in the next seven days, I will test ___ before allowing AI to ___.”

15. Assessment system

BEFORE - baseline capability

Scenario item 1

A team says an AI drafting tool reduced memo time by 40% in three demos. What is the best next step?

A. Deploy to everyone because the time saving is large. B. Compare representative human-only and human-AI cases using a quality rubric and failure criteria. C. Ask users whether they liked it. D. Improve the prompt until the output is longer.

Correct: B. Measures test-design judgment.

Scenario item 2

Which signal most strongly suggests a task needs human-led authority?

A. The task is repetitive. B. The output is easy to reverse. C. Errors are hard to detect and create irreversible external commitments. D. The model answers quickly.

Correct: C. Measures consequence-aware role assignment.

Baseline performance task

Give participants a five-step workflow and ask them to: identify one task to test, name a baseline, choose one metric, and place a human checkpoint. Score with the application rubric below. Retain the same structure, with a new scenario, after the session.

Confidence items

Rate 1-5: “I can decompose a workflow into testable tasks”; “I can design a fair human-versus-human-AI comparison”; “I can decide where final authority belongs.” Confidence is interpreted only alongside performance.

DURING - formative assessment

- Hinge question after research: “If two tasks look equally difficult to a manager, should they receive the same AI policy?” Correct answer: not without task-level evidence.
- Demo poll: “What failed first: model capability, information availability, verification design, or all three?” Debrief accepts multiple causal layers but requires evidence.
- Worksheet checkpoint: facilitator scans for task-level granularity and a measurable baseline.
- Peer challenge: partner must identify one edge case and one way a polished error could survive.

AFTER - demonstration of learning

Application rubric, 0-3 per criterion

Criterion	0	1	2	3
Task decomposition	Job/department named	One broad activity	Multiple tasks but unclear boundaries	4-7 observable tasks with inputs, outputs, decisions, owners
Frontier hypothesis	No rationale	Preference-based rationale	Some task cues	Evidence-based rationale with uncertainty explicitly stated
Test design	Demo only	Unrepresentative case or no baseline	Baseline plus typical cases	Same rubric, credible baseline, 3-5 representative cases including edge case
Authority and guardrail	“Human checks”	Human named but check unclear	Evidence source or escalation named	Accountable human, independent evidence, escalation and stop rule
Metrics and action	Vanity/activity metric	Time or volume only	Quality and time	Quality/correctness, time, failure severity, owner, review date, decision rule

Maximum 15. Evidence of session-level learning: median gain of at least 3 points from baseline and at least 80% of participants scoring 2 or higher on test design and authority. Thresholds are prototype targets and should be calibrated after pilots.

LATER - transfer and impact

Seven-day follow-up

- Did you run the first representative case? yes/no.
- Upload or summarize the test design without confidential data.
- What changed: task boundary, prompt/context, verification, operating mode, or decision to stop?
- Evidence: time, quality, error, override, or no result yet.

Thirty-day follow-up

- Number of tasks tested, moved to controlled pilot, revised, or stopped.
- Median cycle-time change and quality difference versus baseline for tested tasks.
- Critical-error count and near misses.
- Percentage of pilots with a named owner, evidence source, and stop rule.
- One decision changed because evidence contradicted initial enthusiasm.

Ninety-day follow-up

- Sustained usage only for tasks that passed criteria.
- Business outcome linked to the workflow: throughput, rework, conversion, service level, or risk reduction.
- Drift/exception rate and human override pattern.
- Evidence of capability transfer: participants independently apply the loop to a new task.

Meaningful versus vanity metrics

Meaningful: rubric gain, representative-test quality, correctness, error severity, time adjusted for rework, transfer to a new task, controlled experiments completed, decisions revised or stopped, and business outcome versus baseline.

Vanity when isolated: attendance, slides viewed, prompts submitted, number of AI logins, raw output volume, satisfaction, and self-reported confidence. These can diagnose access or experience but do not prove learning or value.

16. Feedback system

Post-session instrument

1. Which step can you now perform more reliably: decompose, predict, test, assign, or monitor? Provide evidence from your worksheet.
2. At what moment were you least sure what to do next? What instruction or example was missing?
3. Which research claim or limitation remained unclear?
4. Was the main exercise too short, about right, or too long? Which part needed different timing?
5. Did your selected workflow produce a concrete next action? yes/no. If no, what blocked it?
6. Identify one example, phrase, visual, or technology requirement that excluded or slowed you.
7. What curriculum change would most improve your ability to run the seven-day test?

Analysis workflow

Participant feedback -> AI-assisted coding -> theme identification -> learning-performance data -> facilitator observations -> recommended revisions -> human review -> curriculum update -> version history.

AI may cluster comments, propose theme labels, connect comments to timestamps, and draft a change summary. A human reviews original comments, checks minority and accessibility signals that clustering may hide, joins them with rubric results, and decides whether the evidence warrants a change.

Example after Cohort 1

Observed: 18 of 24 participants completed a task map, but only 9 wrote a true stop rule. Feedback clustered around uncertainty between “success metric” and “stop rule.” Facilitator notes showed that the demo ran four minutes long, compressing experiment planning. Satisfaction was 4.7/5, but this did not change the diagnosis.

Revision: shorten the demo by two minutes; add a counterexample (“20% faster but one critical unsupported claim = stop”); place the stop-rule field before the operating-mode decision; add a facilitator checkpoint at minute 40. Version changes from 0.8 pilot to 0.9 pilot. Validate the revision by comparing stop-rule rubric scores in Cohort 2.

17. AI-enabled production workflow

Workflow

Faculty research -> controlled research ingestion -> evidence extraction -> claim ledger -> learning-objective candidates -> instructional architecture -> faculty/SME validation -> workshop production -> deck/workbook/guide/exercises/digital assets -> accessibility and factual QA -> delivery readiness -> facilitation -> assessment -> feedback analysis -> revision recommendation -> human approval -> version control and archive.

Human and AI responsibility matrix

Stage	Appropriate AI use	Human responsibility	Faculty approval	Primary risk
Intake and ingestion	Parse provided files; extract metadata; create searchable corpus	Confirm rights, version, source of record, sensitive-data handling	Not always	Wrong or outdated source
Evidence extraction	Produce claim-evidence-limit table with quotations below allowed limits and page references	Read source; verify every material claim and causal verb	Yes for interpretation	Hallucinated or decontextualized finding
Executive implications	Generate multiple cautious translations	Decide relevance and distinguish implication from author conclusion	Yes	Overstatement and false prescription
Objectives and architecture	Suggest measurable verbs, alignments, timing options	Select outcomes and ensure activity/assessment alignment	Review	Attractive activities without learning value
Scenario and exercise	Generate fictional cases, variants, edge cases	Check realism, inclusivity, difficulty, confidentiality, and construct validity	Review when research interpretation is embedded	Scenario accidentally teaches the wrong lesson
Asset production	Draft slide copy, notes, scripts, alt text, formatting variants	Curate, edit, ensure coherence and voice	Review substantive claims	Content sprawl and inconsistency
Assessment	Generate item pools and rubric drafts	Validate that scores measure the objective; calibrate thresholds	Review substantive items	Measuring recall or satisfaction instead of capability
QA	Compare claims to ledger; detect missing sources, contradictions, reading level, accessibility issues	Resolve flags; visually inspect; run files; approve release	Final claim review	False pass from automated checks
Delivery	Produce run-of-show reminders, participation summaries	Facilitate, read the room, protect psychological safety, adapt timing	No	Automation interfering with human interaction
Feedback and revision	Cluster comments, compare cohorts, draft change log	Inspect raw data, protect minority signals, approve changes	For interpretation changes	Majority feedback erasing important edge cases

Automation boundaries

Appropriate: file naming, metadata extraction, draft transformation, cross-asset consistency checks, link validation, version diffs, alt-text drafts, feedback coding, and release manifests.

Inappropriate without human review: deciding what the research means, approving learning objectives, determining ethical acceptability, grading nuanced executive work solely with an LLM, publishing faculty-attributed claims, or autonomously changing validated curriculum.

18. SOP - AI-Assisted Experiential Learning Production

Purpose and scope

Turn validated faculty research into delivery-ready experiential learning while preserving academic integrity, participant usefulness, accessibility, and version control. Applies to workshops, short digital modules, facilitator-led executive experiences, and derivative assets.

Roles

- Faculty/SME: authority on research interpretation and boundaries.
- Instructional designer: owner of learning alignment, production system, and release readiness.
- Facilitator: authority on delivery conditions and observed learner behavior.
- Producer/editor: asset quality, media production, and consistency.
- Program/GTM partner: audience, client context, market-facing accuracy, logistics.
- Data/measurement partner: instrument design, privacy, analysis, and reporting.

Phase 1 - Intake

Checklist:

- Confirm audience, use context, duration, delivery mode, sponsor, and decision date.
- Identify source of record, publication status, version, authors, and usage rights.
- Record requested outputs and channels.
- Classify data and confidentiality.
- Create project ID, owner, folder, decision log, and version 0.1.
- Define faculty review dates before drafting.

Exit criterion: signed intake brief and source manifest.

Phase 2 - Research validation

Checklist:

- Read the full primary source, not only abstracts or summaries.
- Extract research question, method, sample, measures, findings, limitations, and unresolved questions.
- Build claim-evidence-limit ledger with page references.
- Separate author claims, instructional implications, and prototype assumptions.

- Run evidence-verification prompt; manually check every high-consequence claim.
- Mark “requires faculty validation” wherever public evidence is insufficient.

Exit criterion: faculty-ready research brief with no uncited material claim.

Phase 3 - SME collaboration

Send the brief at least three business days before the review. Ask faculty to correct interpretation, identify non-negotiable nuance, flag prohibited extrapolations, and suggest the executive decision the research should inform. Capture decisions, not only comments.

Exit criterion: interpretation approved, approved with changes, or explicitly labeled prototype-only.

Phase 4 - Objectives and design

Checklist:

- Write 3-5 observable objectives using backward design.
- Map knowledge, skill, activity, and assessment for each objective.
- Choose a real-work transfer task.
- Design the session rhythm; no passive block over ten minutes without rationale.
- Select formats only when they serve an objective.
- Define measurement before asset production.

Exit criterion: alignment map approved by instructional-design lead and SME.

Phase 5 - Production

Create a single content source and derive deck, workbook, guide, digital text, and media scripts from it. Use controlled prompts with the source ledger attached. Assign unique IDs to claims, objectives, activities, and assets. Draft in low fidelity first; lock narrative before polishing.

Exit criterion: complete alpha assets with source blocks and no placeholders.

Phase 6 - Review

Review lanes:

1. Faculty/SME - claims, nuance, implications.

2. Instructional design - alignment, cognitive load, transfer, assessment.

3. Facilitator - timing, language, room flow, likely responses.

4. Accessibility - reading order, contrast, alt text, captions, keyboard access, alternatives.

5. Program/GTM - audience and logistics; no implied endorsement or unsupported promise.

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

Phase 7 - QA and release

Checklist:

- Verify every research claim against the ledger and source URL.
- Confirm all “requires validation” labels remain where needed.
- Run spelling, link, consistency, accessibility, and file-integrity checks.
- Render and inspect every page and slide.
- Pilot all instructions, timing, and transitions.
- Test virtual links, permissions, downloads, video, audio, and fallback assets.
- Confirm participant files contain no facilitator-only answers.
- Generate release notes and manifest.

Release naming: `PROJECT_ASSET_AUDIENCE_vMAJOR.MINOR_YYYY-MM-DD.ext`.

Version rules: major = objective/research interpretation or experience redesign; minor = content/activity revision; patch = typo, link, or non-substantive production fix.

Phase 8 - Delivery

At T-48 hours confirm roster, accommodations, tools, files, and facilitator roles. At T-30 minutes test room/virtual setup and open offline fallbacks. During delivery log timing, questions, misconceptions, and deviations without collecting unnecessary personal data.

Phase 9 - Assessment and iteration

Join learner-performance data, feedback, facilitator observation, and transfer evidence. AI may summarize; a human reviews raw evidence and recommends keep/change/test decisions. Every revision cites its evidence, owner, and validation plan.

Phase 10 - Archive

Archive source manifest, approved research brief, final assets, editable sources, release manifest, change log, assessment instruments, de-identified results, approval record, and retirement date. Mark superseded files and preserve the latest approved version as source of truth.

19. Prompt library

Shared safeguard block

Append this block to every research-based prompt:

“Use only the 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 the sample and task. If evidence is missing, write ‘not established in the supplied source.’ Every research claim must include a source and page/section pointer. Output unresolved items under ‘Requires faculty validation.’”

1. Research extraction

“Act as a research analyst supporting an instructional designer. From the supplied 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. Use the shared safeguard block.”

2. Evidence verification

“Audit the draft claim ledger against the primary source. For each claim return supported / partially supported / unsupported / overstated; the exact source location; what wording must change; and whether faculty validation is required. Treat causal verbs, population claims, and numbers as high risk. Use the shared safeguard block.”

3. Executive implications

“Generate five possible executive implications from the validated claim ledger. For each, show the finding it derives from, the reasoning bridge, the boundary conditions, the decision it could inform, and the risk of misinterpretation. Label all implications as proposed translations, not author conclusions, unless explicitly stated by the authors.”

4. Learner analysis

“Given the audience data, create evidence-based learner personas covering role, context, prior fluency, motivation, fear, misconception, task need, accessibility need, value signal, and disengagement risk. Do not fabricate demographic traits. Identify what must be learned before, during, and after the experience.”

5. Learning objectives

“Using backward design, propose 3-5 objectives that state what learners will do in a new situation. For each: required knowledge, required skill, aligned practice, aligned assessment, success criterion, and research claim supporting inclusion. Reject verbs such as understand, know, learn, or appreciate unless paired with observable performance.”

6. Workshop architecture

“Design three 60-minute architectures using Learn -> See -> Do -> Reflect -> Apply. No passive block over ten minutes. Map every segment to an objective and assessment. Include transitions, cognitive-load rationale, and a technology-failure variant. Do not introduce research claims outside the ledger.”

7. Exercise generation

“Create three real-work exercises that require participants to make a decision or build an artifact. For each: objective, instructions, inputs, outputs, time, rubric, facilitator role, common mistakes, accessibility alternative, and transfer step. Explain how the exercise embodies the research without reenacting or copying the study.”

8. Case and scenario generation

“Create a fictional executive scenario with realistic but invented data. Include one decisive piece of qualitative evidence that conflicts with surface-level quantitative evidence. Clearly label every person, company, and number fictional. Produce learner version, facilitator answer key, likely AI error, and alternate valid interpretations. Do not borrow proprietary case content.”

9. Facilitator notes

“For each supplied slide, write SAY, ASK, DO, WATCH FOR, and DEBRIEF notes. Include expected participant responses, transition, time, optional cut, virtual adaptation, and technology fallback. Preserve the approved source block verbatim.”

10. Assessment creation

“For each objective, create one baseline item, one formative check, one post-performance task, and one transfer measure. Provide answer key or analytic rubric and explain construct alignment. Avoid trivia, satisfaction proxies, and items answerable by superficial keyword matching.”

11. Misconception identification

“From the research brief and draft curriculum, list likely misconceptions. For each: why it is attractive, evidence that corrects it, diagnostic question, facilitator response, and activity or visual that prevents it. Do not invent what the authors believe.”

12. Accessibility review

“Audit the content for cognitive load, plain language, reading order, contrast dependency, color-only meaning, alt text, captions/transcripts, keyboard use, motion, screen-reader structure, low bandwidth, language fluency, AI familiarity, and cultural assumptions. Return issue, learner impact, exact fix, owner, and priority.”

13. Content QA

“Compare deck, workbook, facilitator guide, assessment, and digital copy. Flag contradictory numbers, mismatched terminology, missing sources, objective-activity-assessment gaps, facilitator answers exposed to participants, timing conflicts, outdated versions, unsupported promises, and disclaimer inconsistencies. Do not silently fix; produce a release-blocking issue list.”

14. Feedback analysis

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

15. Curriculum revision

“Given approved findings from the feedback analysis, propose the smallest curriculum revisions likely to improve the target objective. For each: problem, evidence, exact asset/location, proposed change, risk introduced, approval required, version increment, and validation measure for the next cohort. Do not modify validated research interpretation without faculty approval.”

20. Multimedia specifications

75-second explainer video

Learning objective served: explain why a workflow must be tested task by task.

Storyboard and script:

1. 0-8 seconds - Visual: one smooth line labeled “what we expect.” Voice: “We often imagine AI capability as a smooth line: easy tasks here, hard tasks there.”
2. 8-20 - Visual: line becomes jagged and crosses a five-step workflow. Voice: “HBS-affiliated research offers a more useful picture. The frontier is jagged. Tasks that look similarly difficult to us can land on opposite sides for AI.”
3. 20-35 - Visual: three evidence figures. Voice: “In a randomized experiment with 758 consultants, AI users completed more work, moved faster, and produced higher-rated work on tasks inside the tested frontier.”
4. 35-48 - Visual: accuracy bars 84.5, 70.6, 60. Voice: “On a different task, accuracy fell. The dangerous part: the wrong answer could still sound coherent.”
5. 48-65 - Visual: Frontier Fit Loop. Voice: “So do not automate a job by intuition. Decompose the workflow. Predict. Test representative cases. Assign human and AI authority. Monitor what changes.”
6. 65-75 - Visual: blank experiment card. Voice: “The goal is not to find the frontier once. It is to build a repeatable way to learn where it is before the risk travels.”

End card disclaimer: independent candidate prototype; research source linked; requires faculty validation.

Infographic information architecture

Title: “One workflow. Two sides of the frontier.” Top: study design in one row. Middle: two vertical lanes - inside and outside the tested frontier - with performance evidence and a “do not generalize” note. Bottom: Frontier Fit Loop with a moving-boundary icon. Footer: primary source, date, disclaimer, accessible text alternative. Use no color-only encoding; pair “helps” and “harms” with symbols and labels.

Framework diagram

A five-step loop with one question per node: Decompose - what is the task? Predict - what is our hypothesis? Test - what does representative evidence show? Assign - who/what holds authority? Monitor - what changed? A moving dashed boundary sits behind the loop, not as a classifier. Each node maps to a workbook page.

Interactive activity

Learner enters a workflow and receives guided questions, not a verdict. The tool converts the workflow into candidate tasks, asks the learner to correct them, records a prediction, generates test-case prompts, asks for acceptance criteria and risks, and exports a human-approved experiment plan. It refuses to classify high-stakes work as safe or to accept confidential data.

Takeaway card

Front: five-step Frontier Fit Loop and operating-mode rule. Back: representative-test checklist, “coherent is not correct” warning, seven-day experiment fields, and re-test triggers. A separate one-page PDF is included in the package.

21. Optional digital experience specification

Product concept - Frontier Fit Companion

Purpose: extend workshop transfer by helping an executive turn a real workflow into a structured, reviewable experiment plan.

User journey and screens

1. Welcome and boundaries - disclaimer, no confidential data, learning tool not approval engine.
2. Define outcome - business outcome, current baseline, owner, consequence of error.
3. Map workflow - enter steps; AI proposes task boundaries; user must edit and approve.
4. Predict - user classifies each task before seeing assistance; records rationale and confidence.
5. Select task - tool suggests a low-risk, reversible starting point; user decides.
6. Build test - typical, edge, and ambiguous cases; baseline; blind rubric; stop rule.
7. Assign mode - AI-led, human-AI, or human-led; accountable human, evidence source, escalation.
8. Review risks - privacy, security, bias, legal, reputational, operational, deskilling, and accessibility prompts.
9. Export - one-page experiment plan, decision log, calendar-ready review date, and unresolved questions.
10. Follow-up - capture result, override, exception, and go/revise/stop decision.

Inputs

Workflow name, desired outcome, tasks, allowed data types, baseline metrics, error consequence, representative cases, evaluation rubric, owner, evidence sources, tool/model configuration, review date.

Outputs

Human-approved workflow map, frontier hypotheses, test protocol, risk register, operating-mode decision, success/stop criteria, seven-day action plan, and versioned result summary.

AI behavior

Ask questions, suggest decomposition, detect vague metrics, propose edge cases, identify missing evidence, and draft the export. It must show its reasoning as questions and assumptions, not issue a safety certification.

Guardrails

- No claim that a task is “safe,” “inside the frontier,” or approved.
- High-stakes categories trigger human-led default language and expert-review prompts.
- Prohibit secrets, personal data, health data, student data, legal matter details, and unreleased strategy unless an approved enterprise environment and policy allow them.
- Every generated statement labeled suggestion, learner input, or evidence.
- Record model/configuration and date because the frontier moves.
- Human confirmation required before export and after every substantive AI rewrite.

System prompt

“You are the Frontier Fit Companion, a learning and experiment-design assistant. Help the user decompose a workflow, state hypotheses, design representative comparisons, assign accountable human and AI roles, and define success and stop criteria. You do not certify safety, compliance, accuracy, or organizational approval. Never infer that a task is inside an AI capability frontier from description alone. Ask for evidence, uncertainty, reversibility, and consequence. Do not request or retain confidential, personal, regulated, or proprietary information. If the task is high stakes, difficult to verify, legally dispositive, medical, employment-related, financial, safety-critical, or irreversible, recommend a human-led design and qualified review. Preserve the user's words separately from your suggestions. Label assumptions. Generate an export only after the user confirms each section.”

Data requirements

Minimal data by default; no raw organizational content required. Store project ID, user-entered abstracted workflow, configuration, decision log, and de-identified outcome metrics. Retention and analytics require organizational policy approval. Provide delete/export controls.

Error states

Vague workflow; sensitive data detected; no baseline; no representative cases; metric only measures speed; no accountable owner; high-stakes task; contradictory success and stop rules; model unavailable; export failure. Each error gives a plain-language recovery path and offline worksheet link.

Evaluation criteria

Task decomposition rubric, completeness of test design, meaningful metric rate, appropriate authority assignment, unresolved-risk visibility, export accuracy, completion without sensitive data, seven-day test initiation, and 30-day transfer to a second workflow.

22. Accessibility and inclusive-design review

Issue	Concrete improvement
Cognitive load	One decision per slide; research numbers split across two slides; five-step framework repeated in consistent order; workbook reduces copying
Readability	Minimum 24-point mid-level slide text and 16-point body; short lines; plain-language definitions; no research paragraphs on slides
Visual accessibility	Contrast checked; color paired with words and symbols; charts include direct labels; no meaning dependent on red/green; alt text for every informative image
Hearing access	Live captions for virtual/in-person amplification; transcript and script for video; facilitator repeats audience questions
Screen-reader and motor access	Tagged heading structure in documents; real table headers; descriptive links; keyboard-operable digital flow; no drag-only activity
AI familiarity	Optional two-minute tool orientation; prepared outputs mean live prompting is not required; glossary defines model, task, baseline, frontier, and guardrail
Technology access	Printable workbook; low-bandwidth HTML; offline exercise; no account required for the core session; facilitator-owned demo
Professional diversity	Examples use common operating decisions; participants apply the framework to their own context; peer pairs cross functions when possible
Cultural assumptions	Avoid idioms, war metaphors, and US-only regulatory shorthand; invite localized constraints and definitions of consequence
Psychological safety	Participants can use a fictional sample; no public disclosure of confidential failures; critique the workflow, not the person
Language complexity	Define “jagged frontier” in one sentence; use task-level verbs; provide a glossary and written instructions before timed work
Risk and inclusion	Explicitly ask who bears the cost of error and whose context is missing from training/test cases; include accessibility as a risk dimension

23. Portfolio case study

The challenge

How can complex AI research become something an executive can understand, practice, and apply within one hour?

The research

The selected peer-reviewed study tested how 758 consultants performed with and without GPT-4 on realistic knowledge-work tasks. AI improved completion, speed, and quality on one set of tasks and reduced correctness on another, motivating the jagged-frontier concept. The result is not a universal capability map. It is evidence that the effect of AI can reverse within a workflow. Primary paper.

The translation

Research question -> executive problem: “How should I decide where AI gets authority in a workflow when a polished output is not proof of fit?”

Learning objectives

Participants learn to decompose, hypothesize, test, assign, and monitor one task. The assessment asks them to build the decision artifact, not recall a definition.

The experience

The session opens with a deployment decision, teaches only the research needed for action, demonstrates a persuasive failure, and gives most of the remaining time to a real-work Frontier Fit Map and seven-day experiment.

The exercise

Participants do not practice prompting as the end goal. They create a representative test, evidence rubric, human checkpoint, stop rule, and transfer plan.

Artifacts

Facilitation deck; participant workbook; completed example; live demo; facilitator guide; assessment and feedback systems; AI-enabled production workflow; SOP; prompt library; multimedia specifications; digital-companion specification; accessibility review; executive card; microsite copy; interview deck.

Measurement

Learning is measured with a pre/post application rubric. Transfer is measured at 7, 30, and 90 days through experiments run, decision quality, error severity, business outcomes, and independent use on a new workflow.

The system

AI accelerates evidence extraction, draft transformation, scenario variation, consistency checks, accessibility review, feedback coding, and change summaries. Humans retain research interpretation, learning design, ethical judgment, facilitation, release approval, and revision authority.

Iteration

Each revision requires evidence, a named owner, a version increment, an approval route, and a validation measure. Satisfaction can surface friction but cannot override performance data.

What I would validate with HBS faculty

- Whether the Faculty Research Brief preserves the authors' intended interpretation and boundary conditions.

- Whether “Frontier Fit Loop” is a useful translation or risks implying the frontier can be located too confidently.
- Whether the executive implications are appropriate for the Institute's audience and current research portfolio.
- Whether the demo captures overreliance without oversimplifying the mechanism in the study.
- Which terms, examples, or metrics HBS already uses and should remain consistent.
- What research claims require updated figures, author-preferred wording, or removal.
- How the Institute distinguishes enterprise learning, executive education, and public digital learning in its internal design standards.

Everything beyond the published findings is an independent candidate prototype. Nothing here implies Harvard review, endorsement, or authorship.

24. Portfolio microsite structure and copy

Global design

Academic, modern, restrained. Warm white background, charcoal text, muted indigo accent, one serif display face paired with an accessible sans serif, generous whitespace, direct labels, no shields, crimson imitation, Harvard wordmarks, or visual suggestion of endorsement. Sticky section navigation on desktop; linear reading on mobile. Reading time: five-minute overview with optional deep dives.

Persistent disclaimer: “Independent candidate prototype created using publicly available HBS research. Not an official Harvard Business School product.”

Hero

Eyebrow: Independent instructional-design proof of work

Title: From HBS Research to Experiential AI Learning

Copy: “A research-to-learning prototype that turns one peer-reviewed study into a 60-minute executive workshop, a real-work decision tool, a measurement system, and a repeatable AI-assisted production workflow.”

Primary action: Explore the 5-minute overview. Secondary action: Open artifacts.

Proof strip: “1 peer-reviewed study / 60-minute experience / 24-slide workshop / 4 levels of measurement / 15 reusable prompts.”

Research

Headline: The effect of AI can reverse from one task to the next.

Copy: “In a randomized experiment with 758 consultants, AI improved speed, completion, and quality on tasks inside the tested capability frontier. On a different task, accuracy fell. The insight is not that AI is good or bad. It is that workflow design needs evidence at the task level.”

Links: Read the faculty brief; open primary source; see limitations.

Learning design

Headline: The learning outcome is a decision, not a definition.

Copy: “Participants learn to decompose a workflow, predict where AI may help or harm, test representative cases, assign human and AI authority, and monitor what changes.”

Visual: objective -> activity -> evidence table.

Experience

Headline: Learn. See. Do. Reflect. Apply.

Copy: “The session alternates research, demonstration, real-work practice, peer challenge, and transfer. No passive block lasts more than seven minutes.”

Timeline: provocation 5 / research 7 / framework 8 / demo 10 / exercise 15 / debrief 8 / action 7.

Exercise

Headline: Map one workflow before you automate it.

Copy: “The Frontier Fit Map converts a broad AI idea into a representative comparison, a human checkpoint, a stop rule, and a seven-day experiment.”

CTA: View blank worksheet / View completed example.

Facilitation

Headline: Designed to be delivered by someone else.

Copy: “The guide includes minute-by-minute direction, suggested language, expected responses, misconceptions, difficult questions, accessibility accommodations, timing variants, and a no-technology fallback.”

Measurement

Headline: Satisfaction is feedback. Performance is evidence.

Copy: “A pre/post application rubric measures decomposition, test quality, authority design, guardrails, and metrics. Seven-, thirty-, and ninety-day follow-ups look for transfer and organizational impact.”

AI system

Headline: AI accelerates production. Humans remain accountable.

Copy: “AI can extract, transform, compare, check, and cluster. Faculty validate research interpretation. Instructional designers own alignment. Facilitators own the room. Humans approve every release and revision.”

Visual: production workflow with approval gates.

Artifacts

Cards: Faculty research brief; workshop deck; participant workbook; experiential exercise; facilitator guide; assessment system; production SOP; prompt library; multimedia; digital companion; executive card; interview presentation.

Closing

Headline: What I would validate next

Copy: “This prototype uses public evidence and explicitly marks its assumptions. The next step would be faculty and team review: correct the research translation, test the exercise with executives, calibrate the assessment, and revise from evidence.”

Final line: “This is how I approached the problem with the information publicly available to me. I would be excited to understand how the HBS AI Institute approaches it internally.”

25. Three-minute walkthrough

“One thing that stood out to me in the role was the challenge of translating emerging AI research into experiential learning. Rather than only explain how I would approach that, I wanted to test the process myself.

I began with the live job description and treated it as a product specification: research translation, facilitator-ready assets, measurement, AI-enabled production, QA, and continuous improvement all had to show up in the work.

I reviewed nine pieces of public HBS and HBS-affiliated AI research. I selected *Navigating the Jagged Technological Frontier* because it combines rigorous evidence with a decision leaders actually face. In the experiment, AI helped consultants complete more work, move faster, and produce higher-rated work on one set of tasks. On another task, accuracy fell. The durable insight is that one workflow can cross the capability frontier several times, and the boundary is difficult to see in advance.

I translated that into one performance outcome: after an hour, a participant should be able to take a real workflow, decompose it into testable tasks, design a representative comparison, and decide where human judgment belongs.

The workshop moves from a provocation into a concise research brief, then a live demonstration where a polished AI recommendation misses decisive evidence. Participants then use the Frontier Fit Loop - Decompose, Predict, Test, Assign, Monitor - on their own work. They leave with a seven-day experiment, not just notes.

To measure learning, I created a pre/post application rubric and seven-, thirty-, and ninety-day transfer checks. The meaningful question is not whether participants liked the session or used more prompts. It is whether they can design a better test, prevent consequential errors, and improve an actual workflow.

I also built the production system behind the experience: a faculty research brief, claim ledger, faculty approval gates, deck, workbook, facilitator guide, assessment, feedback loop, SOP, and a prompt library with safeguards against fabricated citations and overstated causality. AI accelerates extraction, transformation, consistency checks, and feedback analysis. Humans retain research interpretation, learning design, ethical judgment, facilitation, and release approval.

This is an independent candidate prototype, not an HBS product. The framework and executive implications are my translation of public evidence, and I have marked what requires faculty validation. This is how I approached the problem with the information publicly available to me, and I'd be excited to understand how your team approaches it internally."

26. Five-to-seven-minute interview presentation

Slide 1 - I tested the core challenge in the role

On-screen: "Can public HBS AI research become an executive experience that changes a real decision?" Notes: open with curiosity and the independent-prototype disclaimer.

Slide 2 - I used the job description as the specification

On-screen: Research translation / experiential assets / delivery readiness / measurement / AI production / continuous improvement. Notes: show the HBS Role -> Evidence Map, not a generic design process.

Slide 3 - I selected a study with a useful tension

On-screen: "Inside the tested frontier: more, faster, higher rated. Outside: less correct." Notes: explain design and limitations in under 60 seconds; link primary source.

Slide 4 - I designed backward from a capability

On-screen: "Learner can test one task and assign AI an evidence-based role." Visual: objective -> demo -> Frontier Fit Map -> rubric -> seven-day experiment.

Slide 5 - The participant does the important work

On-screen: Decompose -> Predict -> Test -> Assign -> Monitor. Notes: walk through the real-work exercise and completed example; label the framework as candidate-created and requiring validation.

Slide 6 - AI accelerated production; humans own judgment

On-screen two columns. AI: extract, transform, compare, check, cluster. Humans/faculty: interpret, select, facilitate, approve, revise. Notes: show where hallucination, privacy, and evidence risks are controlled.

Slide 7 - I would validate, pilot, measure, and revise

On-screen: Faculty accuracy -> executive pilot -> performance rubric -> 7/30/90-day transfer -> versioned revision. Closing: "I built enough to make my thinking testable, not to presume this is how HBS should do it."

27. HBS Role -> Evidence Map

Actual role responsibility or qualification	Evidence in this project
Connector among faculty, SMEs, facilitators, and go-to-market teams	Faculty brief; approval gates; facilitator guide; microsite and market-facing disclaimer
End-to-end lifecycle of experiential learning creation	Research intake through archive SOP and production workflow
Rapid conversion into client-ready and market-facing assets	Cross-format asset family derived from one validated content source
Quality, coherence, consistency, version control, QA, release documentation	Claim ledger, source blocks, QA checklist, naming/version rules, release manifest model
Refresh and maintain facilitation materials, exercises, and learning assets	Editable workshop deck, workbook, exercise, example, guide, and revision loop
Translate complex academic research into accessible participant-centered formats	Faculty research brief -> executive problem -> Frontier Fit Loop -> real-work exercise
Decks, toolkits, videos, blog posts, infographics, and digital content	Deck, workbook/toolkit, video script, infographic IA, case-study/microsite copy, digital companion
Session preparation, facilitator support, asset readiness, post-session refinement	Setup, timing, scripts, fallbacks, readiness checklist, cohort revision example
Participant and facilitator feedback for continuous improvement	Seven-question feedback instrument and human-reviewed AI analysis pipeline
Metrics for learning performance, engagement, and impact	Baseline, formative checks, post rubric, 7/30/90-day transfer measures
SOPs for AI-enabled production and delivery	Complete repeatable production SOP with checklists and approval gates
Virtual and in-person program support	Multi-mode setup, accessibility plan, low-tech fallback, prepared demo outputs
Technical translation for non-technical audiences	Plain-language mental model, glossary, demonstration, difficult-question responses
Responsible AI tool selection and configuration	Human/AI responsibility matrix, source-bounded prompts, privacy rules, digital guardrails
AI-driven workflows, automations, and agents	Production workflow, prompt library, optional companion system prompt, feedback automation design
Research translation for executive audiences	Entire case study and workshop outcome
Working knowledge of AI/ML and agentic systems	Model-version limits, evaluation design, retrieval/fine-tuning boundaries, tool/agent governance in guide and system spec

28. Final quality audit

Harvard professor lens - research accuracy

Initial weakness: the first framework draft risked implying that a task can be permanently classified as inside or outside the frontier.

Fix: changed classification to a hypothesis; inserted representative testing, configuration/date capture, monitoring, and a repeated statement that the framework is a candidate-created translation requiring faculty

validation. Used the peer-reviewed 2026 paper as source of record and separated study findings from executive implications.

Remaining validation: author-preferred wording, interpretation of the prompt-overview condition, and whether the operating-mode labels create unintended claims.

Executive participant lens - usefulness

Initial weakness: a research-heavy session could become intellectually interesting but operationally inert.

Fix: limited research instruction, added a persuasive-failure demo, allocated the largest block to a real workflow, and required a seven-day experiment with owner, metric, stop rule, and review date. Added a completed B2B escalation example.

Remaining validation: pilot timing with a mixed-seniority executive cohort.

Instructional-design leader lens - alignment

Initial weakness: “understand the frontier” was too vague and the feedback survey overemphasized confidence.

Fix: rewrote objectives as observable performance, aligned each to practice and assessment, created a 15-point pre/post rubric, and treated confidence and satisfaction as diagnostic rather than proof. Added transfer measures.

Remaining validation: calibrate rubric thresholds and inter-rater agreement after pilot scoring.

AI expert lens - intelligent use

Initial weakness: a prompt library can become performative if it does not control evidence risk.

Fix: added source-bounded safeguards, claim typing, faculty approval, privacy limits, human review, and non-automation boundaries. The learning objective is workflow evaluation, not prompting.

Remaining validation: align tool selection and data controls with HBS-approved environments and policies.

Hiring-manager lens - proof of capability

Initial weakness: a single polished deck would not prove operational readiness, measurement, or scale.

Fix: built the connected system - job map, research landscape, brief, deck, workbook, exercise, facilitator guide, assessment, feedback loop, AI workflow, SOP, prompt library, multimedia, digital continuation, accessibility review, case study, microsite copy, walkthrough, interview deck, and role-evidence map.

Remaining validation: discuss which artifacts best mirror the team's current priorities and where this prototype should be deliberately smaller.

Release decision

Candidate prototype status: ready for interview review after final file-level visual QA. Not ready for institutional delivery until faculty/SME validation, pilot testing, accessibility review in the target platform, and confirmation of HBS data/tool policies.

Source ledger

- Harvard Careers - Instructional Designer, HBS AI Institute
- HBS AI Institute - About Us
- HBS AI Institute - Accelerate with AI
- HBS AI Institute - Future Proof with AI
- Navigating the Jagged Technological Frontier - peer-reviewed paper
- HBS AI Institute - Back to the Beginnings of AI at Work
- HBS AI Institute - The Cybernetic Teammate
- HBS AI Institute - Everyone Has AI. Which Firms are Going to Win?
- HBS AI Institute - The Real AI Divide Isn't Who Uses It Most
- HBS AI Institute - The Future of Decision-Making
- HBS AI Institute - The Three Ways Professionals Work with AI
- HBS Working Paper - Innovating with Generative AI
- HBS AI Institute - Idea Flow + The Machine
- HBS Faculty - Novice Risk Work
- HBS AI Institute - Competing in the Dark