



USING AI TO EVALUATE RFP RESPONSES

Without Losing the Human Touch

Janette Roush & Skylar Clark | April 7, 2026

Agents of Change Monthly Webinar

THE RFP EVALUATION PROBLEM

- Reading 5-10 proposals of 25+ pages each takes **days**
- Each committee member develops their own system, creating **inconsistent evaluation**
- Compliance details get missed when reviewers are fatigued
- Hard to do apples-to-apples comparison across different proposal formats

SOUND FAMILIAR?

"Three people evaluated the same proposal and scored it a 2, a 4, and a 3 — using the same scorecard."

"We didn't catch that the vendor used another client's name until the finalist interview."

"The budget was \$8,000 over our limit and nobody noticed."

AI AS YOUR EVALUATION CO-PILOT

SYSTEMATIC

Same prompt, same criteria,
same rigor applied to every
single proposal

TRANSPARENT

Shared workspace where
everyone sees the same
analysis — no black boxes

HUMAN-LED

AI reads and organizes.
Humans discuss, debate, and
decide.

AI creates a **level playing field** for every proposal —
then humans bring the judgment that matters.

MEET BEIGE COUNTY

For today's demo, we created a fictional DMO scenario to walk through the full workflow.

THE RFP

Greater Beige County CVB seeks a **\$350,000 integrated destination marketing campaign** — brand, digital, content, social, web, PR, and measurement.

Current slogan: *"Beige County: We're Here Too"*

Independent research called it: "something a ghost would say."

THE RESPONSES

8 agencies submitted proposals — each with a distinct profile:

Strong contenders

Dark horse

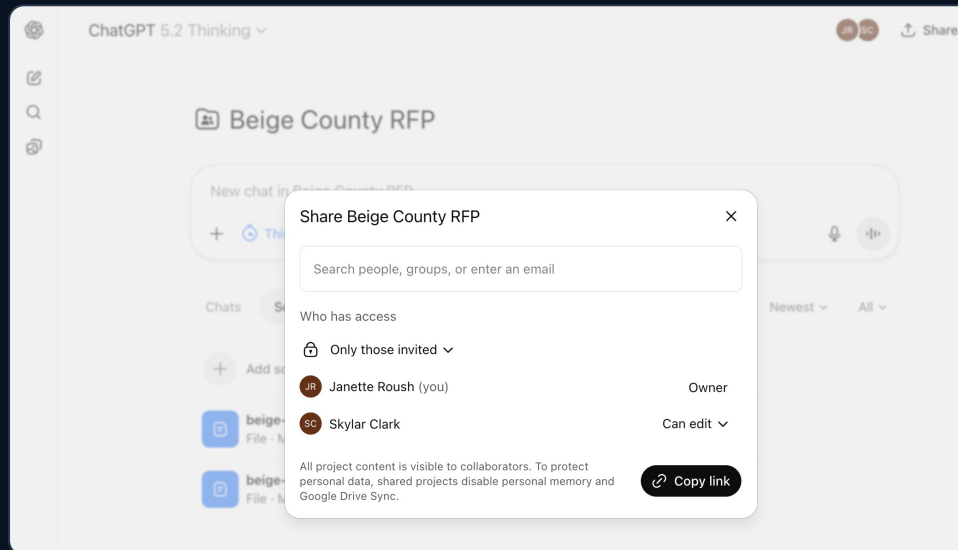
Style over substance

Compliance failures

Plus a detailed evaluation scorecard with 9 weighted criteria sections.

01

SHARE ONE AI WORKSPACE



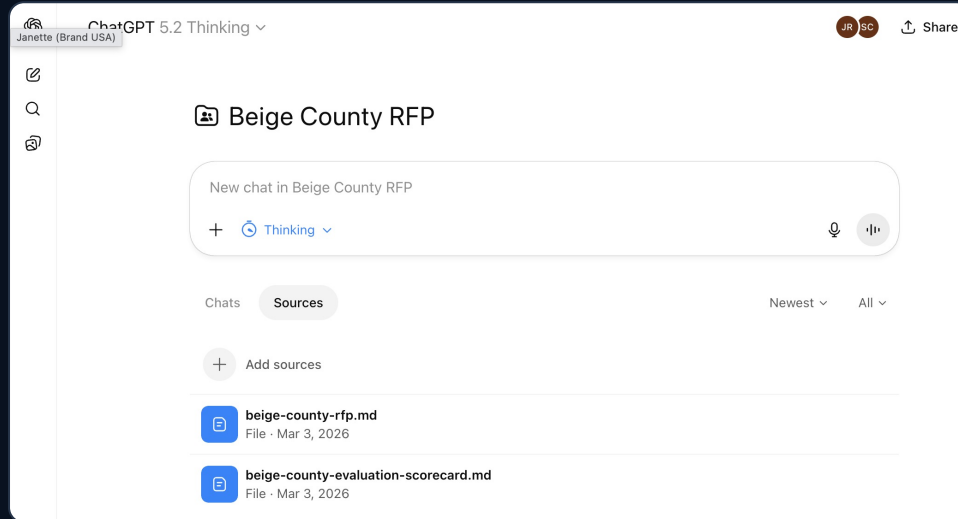
TRANSPARENCY + ZERO DUPLICATE WORK

Give everyone on the RFP committee access to a single shared AI project. Everyone sees the same analysis, the same documents, the same scoring.

- Committee members share **one project** — no one duplicates the setup
- Every conversation is visible — full transparency on how AI was used
- Roles control who can edit vs. view

The alternative: 5 people each uploading files, writing their own prompts, getting different results, and spending hours on setup instead of evaluation.

02 UPLOAD THE SOURCE DOCUMENTS



PROJECT KNOWLEDGE = YOUR GROUND TRUTH

Upload the RFP and evaluation scorecard as **project knowledge**. Every conversation in this project will automatically reference these documents.

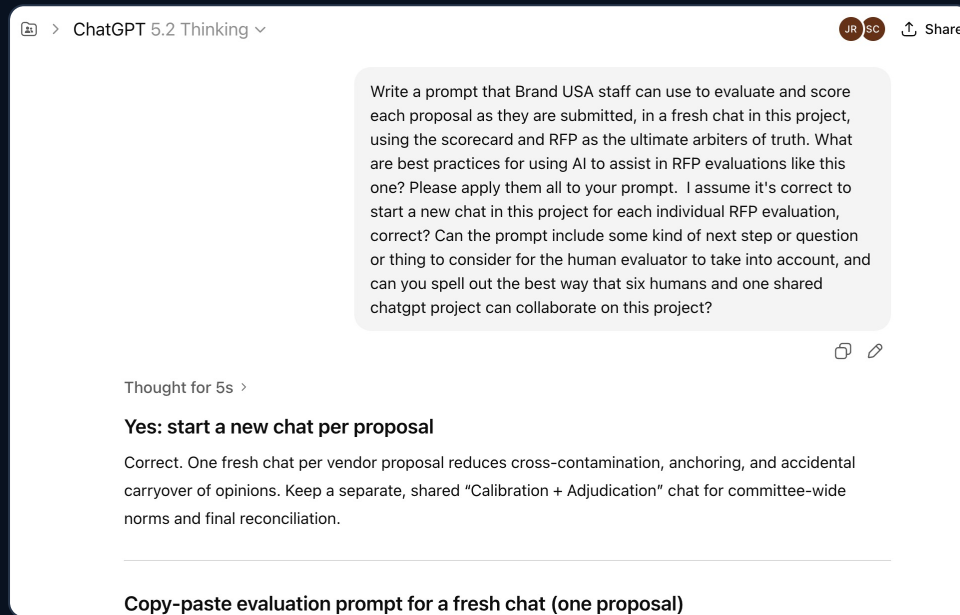
- **The RFP** — what we asked for
- **The evaluation scorecard** — how we're grading

TOOL OPTIONS

ChatGPT Projects — upload files, share with team

Claude Projects — same concept

CREATE A MASTER EVALUATION PROMPT



Ask the AI to help you write the evaluation prompt. It already has the RFP and scorecard — let it do the heavy lifting.

```
"Write an evaluation prompt that staff can use to score each proposal. Use the scorecard and RFP as the source of truth. Walk through each section, assign a score, explain why, flag compliance issues, and note strengths and weaknesses."
```

KEY PRINCIPLES

- **One prompt for all proposals** — ensures consistency
- **Evaluate separately** — one conversation per proposal for a fresh read

STEP 03 CONTINUED

AI GENERATES THE EVALUATION FRAMEWORK

PROMPT (paste into a new chat):

You are an RFP evaluation assistant. Your job is to help a human evaluator score a single vendor proposal using the Greater Beige County CVB Integrated Destination Marketing Campaign RFP and the official Evaluation Scorecard as the only sources of truth. If anything conflicts, the RFP and Scorecard win. Do not invent facts. Do not assume what the vendor meant. If the proposal omits information, mark it as missing and reflect that in scoring.

Authoritative documents:

- Evaluation Scorecard (weights, sub-criteria, compliance checklist, red flags).
 beige-county-evaluation-scoreca...
- RFP (scope A–H, requirements, budget, timeline, compliance requirements, evaluation process).
 beige-county-rfp

Inputs I will provide in this chat:

1. The vendor proposal text (and any appendices) or a summary excerpted from the proposal.
2. (Optional) Notes from my first read.

Operating rules:

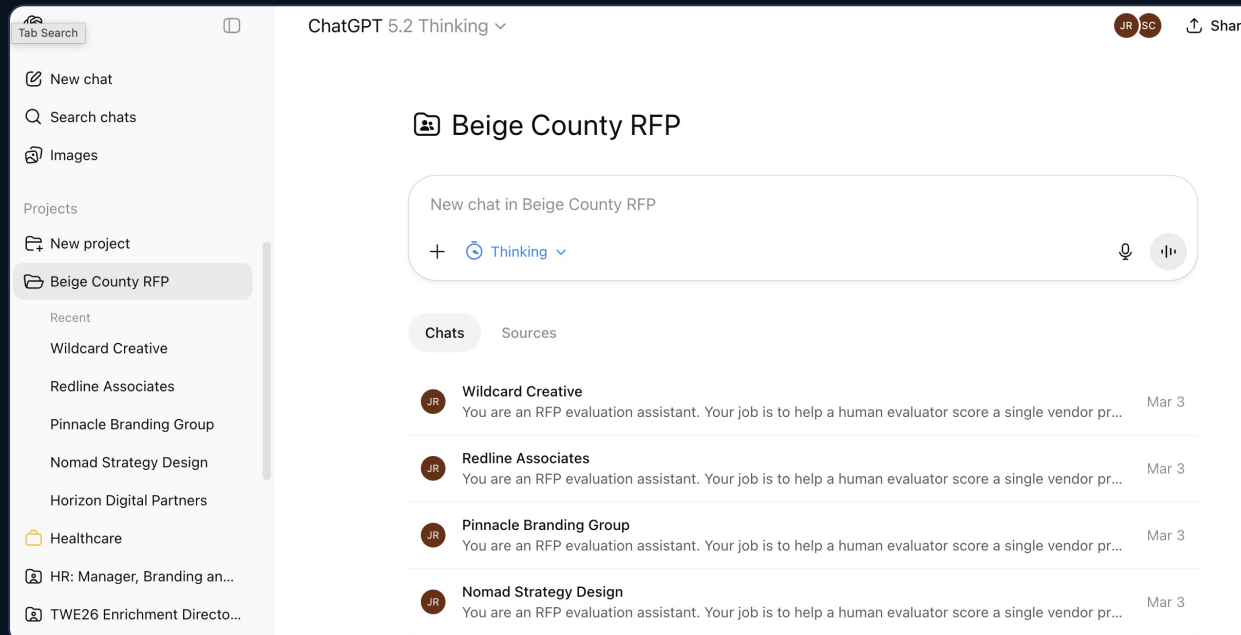
- Evidence-first: every score must cite proposal evidence by quoting short snippets (max ~25 words) and/or giving section/page references as provided.
- No “filling in”: if a detail is not explicitly present  at as not provided.
- Separate compliance from quality: do the Compliance Checklist first, then scoring.

The AI produces a structured evaluation prompt with operating rules, section-by-section scoring criteria, and output format — all derived from your scorecard.

Copy this prompt. You'll paste it into each vendor's evaluation conversation to ensure every proposal gets the same rigorous analysis.

Use the stronger model for this step — ChatGPT Pro, Claude Opus — because depth of reasoning matters when building the framework.

EVALUATE EACH RESPONSE SEPARATELY



Start a **new conversation** for each vendor within the same project. Upload one proposal, paste the master prompt, let the AI work.

- Each vendor gets an **isolated, unbiased evaluation**
- Committee members can review any vendor's thread at any time
- The sidebar becomes your evaluation dashboard

With 8 proposals, this saves 15-20 hours of committee reading

INSIDE EACH EVALUATION

COMPLIANCE CHECK & RED FLAGS

Compliance			
Requirement (Scorecard)	Met? (Y/N/Unclear)	Proposal evidence (quote + section/lines)	Impact
Proposal does not exceed 25 pages (excl. appendices & compliance docs)	Unclear	Proposal provided as text; no page count stated.	Score deduction likely (admin risk)
Includes agency overview	Y	"Nomad Strategy + Design is a place-based storytelling studio..." response-nomad-strategy-design	Proceed
Minimum 3 relevant case studies with measurable outcomes	Y	3 case studies with results like "Social engagement increased 580%..." response-nomad-strategy-design	Proceed
At least 1 case study with destination under 100K population	Y	Flamingo Heights "community of 480 residents"; Phoenicia "pop. 309." response-nomad-strategy-design response-nomad-strategy-design	Proceed
Proposed approach addresses all scope areas (A through H)	Unclear	Covers brand, social, content, some website, some PR, reporting; but no paid search plan (RFP requires it) beige-county-rfp and only "Media outreach (2 targeted pitches. ↓ s minimum 3 response-nomad-strategy-design ; no explicit stakeholder business training session (required)	Score deduction likely; could trigger non-compliance discussion

The AI walks through every requirement in your RFP and checks for compliance — page limits, budget caps, required references, conflict of interest disclosures.

Wrong client name

Over page limit

Budget math errors

Missing references

Vague methodology

Strong dark horse

AI catches what tired eyes miss. Redline's 6 compliance failures and Nomad's \$12K budget overrun? Found in seconds.

INSIDE EACH EVALUATION

SECTION-BY-SECTION SCORING

Scoring by Section

1) Relevant Experience & Case Studies (20%)

Sub-criteria scores (1–5) with evidence

- **Number/relevance of destination case studies: 5**
 - 3 place/destination case studies provided response-nomad-strategy-design
- **Measurable outcomes: 4**
 - "Social engagement increased 580%..." response-nomad-strategy-design
 - "89,000 views... 24% uptick..." response-nomad-strategy-design
 - Gap: limited direct visitation, room nights, or spend impact metrics.
- **Small-to-mid-sized destination experience: 5**
 - "community of 480 residents" response-nomad-strategy-design; "Phoenicia (pop. 309)" response-nomad-strategy-design
- **Understanding of DMO/CVB constraints: 3**
 - Some stakeholder sessions noted response-nomad-strategy-design
 - Gap: little explicit discussion of CVB realities (board governance, small staff workload) as framed in Scorecard "Consider." beige-county-evaluation-scorecard
- **Breadth of services demonstrated: 4**
 - Brand + content + paid social/programmatic + reporting response-nomad-strategy-design

Section average: $(5 + 4 + 5 + 3 + 4) / 5 = 4.20$

2) Proposed Approach & Methodology (20%)

- **Clarity/logic of process: 4**
 - Clear 4-movement plan spanning kickoff through 12 months response-nomad-strategy-design
- **Quality of discovery/research: 5**
 - "Stakeholder Conversations... Digital Archaeology... Cultural Cartography" response-nomad-strategy-design
- **Timeline realism (16-week dev + 12-month exec): 5**
 - Development explicitly mapped to Weeks 1–16; execution Months 1–12 response-nomad-strategy-design
- **Stakeholder engagement (board + community + business training): 3**
 - Board/community are present ("Community Story Circle"; stakeholder sessions) response-nomad-strategy-design
 - Gap: RFP requires "One stakeholder training session... for local businesses" beige-county-rfp; not explicitly addressed.
- **Data/insights vs pure creative: 4**
 - "Digital Archaeology... search behavior... Google autocomplete" response-nomad-strategy-design
 - Gap: limited explanation of how findings translate into channel plans/KPIs.
- **Comprehensiveness across scope A–H: 3**
 - Gap: Paid search requirement present in RFP beige-county-rfp but not described in proposal; PR pitch minimum 3 beige-county-rfp vs "2 to 3 pitches" response-nomad-strategy-design.

Section average: $(4 + 5 + 5 + 3 + 4 + 3) / 6 = 4.00$

The AI walks through each scorecard section, assigns a score on your scale, and **explains its reasoning** with specific evidence from the proposal.

■ 9 weighted sections scored individually

■ Evidence cited for every score

■ Strengths and weaknesses identified

■ Consistent framework across all 8 vendors

INSIDE EACH EVALUATION

SCORE SUMMARY & RECOMMENDATION

Score Summary

Section	Weight	Score (/5)	Weighted Score
1. Relevant Experience & Case Studies	20%	4.20	0.84
2. Proposed Approach & Methodology	20%	4.00	0.80
3. Creative Capability	15%	4.17	0.63
4. Digital Strategy & Technical Capability	15%	3.17	0.48
5. Team Qualifications & Fit	10%	3.80	0.38
6. Content & Media Production Plan	10%	3.80	0.38
7. Value & Budget Allocation	5%	3.80	0.19
8. Measurement & Reporting Approach	3%	3.50	0.11
9. References	2%	2.25	0.05
TOTAL	100%		3.84 / 5.00

Each evaluation ends with a weighted score summary across all 9 sections, a recommendation, and — critically — **suggested next steps for the human evaluator**.

The AI tells you what to look for when you open the actual proposal. You spend your time on substance and judgment calls — the reading is already done.

INSIDE EACH EVALUATION

AI PREPARES YOU FOR THE HUMAN WORK

Human Next Steps (5–8 targeted checks)

1. **Reference compliance:** Can they provide at least one additional destination/tourism reference (and dates) to meet the RFP requirement? [beige-county-rfp](#)
2. **Paid search scope:** Where is the **Google Ads (paid search)** strategy, budget, and management approach? [beige-county-rfp](#)
3. **SEO roadmap:** Do they have a real **SEO audit + optimization roadmap** beyond copy refresh language? [beige-county-rfp](#) [response-nomad-strategy-design](#)
4. **Stakeholder training:** How will they deliver the required **local business stakeholder training session** (format, curriculum, materials)? [beige-county-rfp](#)
5. **PR deliverables:** Can they commit to the RFP minimum **3 proactive media pitches** (they currently budget 2)? [beige-county-rfp](#) [response-nomad-strategy-design](#)
6. **Subcontractor vetting:** Who exactly are the **photographer and production crew**, what are their credentials, and are they available in Beige County's timeline? [response-nomad-strategy-design](#)
7. **KPI specificity:** What draft KPI targets would they propose that map to **2.0 nights and \$28M spend** (even if refined after baseline)? [beige-county-evaluation-scoreca...](#) [beige-county-rfp](#)
8. **Submission admin:** Confirm **page limit, PDF format, and on-time submission** (not evidenced in the proposal text provided). [beige-county-evaluation-s...](#) [a...](#) [beige-county-rfp](#)

Interview Questions (if shortlisted)

1. **(Section 9: References)** Which two destination/tourism clients from the last 3 years can you add as references, and what outcomes did you deliver? [beige-county-rfp](#)
2. **(Section 4: Digital)** Walk us through your **paid search (Google Ads)** plan: campaign structure, budgets, keywords/themes, and landing page strategy. [beige-county-rfp](#)
3. **(Section 4: Digital)** Provide a concrete **SEO audit and roadmap**: technical fixes, content strategy, and measurement approach for visitbeigeconomy.org. [beige-county-rfp](#)
4. **(Section 2: Approach)** Map your deliverables explicitly to **Scope A–H** and call out what you are not doing (if anything). [beige-county-rfp](#)
5. **(Section 2: Stakeholders)** Describe the **local business training session**: agenda, materials, who leads it, and how you'll drive attendance. [beige-county-rfp](#)
6. **(Section 2/6: Content)** Your plan includes 2 shoot days for hero films and seasonal capture. How will you ensure coverage across communities/assets with that schedule? [response-nomad-strategy-design](#)
7. **(Section 5: Team/Subs)** Who are your named subcontractors for photo and production, and can you share 2–3 relevant samples and bios? [response-nomad-strategy-design](#)
8. **(Section 8: Measurement)** Show a sample “Monthly Creative Pulse” and explain how it rolls up to CVB goals like **2.0 nights and \$28M spend**. [response-nomad-strategy-design](#) [beige-county-rfp](#)

Each evaluation includes **targeted checks for humans** (what to verify manually) and **vendor-specific interview questions** (if the vendor is shortlisted). AI does the preparation — humans bring the judgment.

05

COMPARE ACROSS ALL RESPONSES

Once all proposals are evaluated individually, ask the AI to compare them.

VENDOR	PROFILE	AI FLAGS
Pinnacle Branding	Polished, experienced, thorough	Strong contender
Wildcard Creative	Creative genius, partner model for digital	Strong contender
Copper & Sage	Deep DMO expertise, budget-conscious	Dark horse
Atlas Tourism	Strategy expert, subcontracts creative	Split capability
Horizon Digital	Data-driven, weak on creative/culture fit	Wrong fit
BlueWave Marketing	Meets all requirements, zero spark	Adequate middle
Nomad Strategy	Beautiful aesthetic, thin methodology	Budget error (\$362K)
Redline & Assoc.	Big agency, copy-paste proposal	6 compliance failures

AI-GENERATED COMPLIANCE CHECK

Ask the AI to build a compliance matrix across all 8 proposals against your requirements checklist.

REQUIREMENT	PINNACLE	WILDCARD	COPPER&SAGE	HORIZON	BLUEWAVE	NOMAD	REDLINE	ATLAS
Under 25 pages	✓	✓	✓	✓	✓	✓	✗	✓
3+ case studies	✓	✓	✓	✓	✓	✓	✓	✓
Budget ≤ \$350K	✓	✓	✓	✓	✓	✗	✗	✓
3 references (2 tourism)	✓	✓	✓	✓	✓	✗	✗	✓
Timeline ≤ 16 weeks	✓	✓	✓	✓	✓	✓	✗	✓
Conflict of interest disclosure	✓	✓	✓	✓	✓	✓	✗	✓
Correct client name throughout	✓	✓	✓	✓	✓	✓	✗	✓

Redline's 6 failures and Nomad's budget math error would take hours for a human committee to compile manually. AI builds this matrix in seconds.

06

COLLABORATIVE SCORING

THE ARTIFACT APPROACH

Build an interactive scoring tool as a Claude Artifact (or similar) that serves as a **shared scorecard for the entire committee**.

- Each committee member enters their scores
- All scores visible in one place
- Real-time aggregate scores and comparisons
- Export to CSV/Excel for documentation
- Revision history — change your score if you reconsider

BEFORE

5 committee members → 5 individual Excel sheets → 1 admin manually compiling scores → summary spreadsheet nobody trusts

AFTER

1 shared artifact → everyone scores in the same place → automatic aggregation → export when ready

Time saved: hours of admin work per RFP cycle

THE HUMAN SEAT

WHERE AI HELPS

- Reading and summarizing 200+ pages of proposals
- Checking compliance against requirements
- Catching budget math errors
- Identifying copy-paste artifacts from other clients
- Creating consistent apples-to-apples comparisons
- Building scoring tools and dashboards

WHERE HUMANS DECIDE

- Cultural fit — will this agency "get" us?
- Chemistry and trust — do we want to work with these people?
- Strategic judgment — which approach is right for our moment?
- Risk tolerance — partner model vs. full-service?
- Gut check — when AI says "adequate" but something feels off
- The final decision — always human

MAKING IT REPEATABLE

ONCE YOU BUILD THE SYSTEM...

- **Save the evaluation prompt** as a template
- **Save the scoring artifact** and customize for each new RFP
- **New RFP?** Swap in the new documents, adjust the scorecard, and the system is ready
- **Assign one person** to set up the project — everyone benefits from that investment

SETUP TIME VS. IMPACT

First RFP: 2-3 hours to build the project, prompt, and scoring tool

Every RFP after that: 30 minutes to customize

Time saved per RFP cycle: 15-25 hours across the committee

Plus: better decisions from consistent, thorough evaluation

BEYOND RFPS

This same workflow applies wherever you need to evaluate multiple options against consistent criteria.

HIRING

Upload the job description as project knowledge. Evaluate interview transcripts against the same criteria. Remove names to reduce bias. Build a shared scoring tool for hiring committees.

VENDOR REVIEWS

Annual vendor performance reviews. Upload the contract and KPIs as project knowledge. Evaluate each vendor's performance report against the same framework.

GRANT APPLICATIONS

Reviewing grant applications or co-op marketing requests. Upload the criteria and guidelines. Evaluate each application consistently.

The pattern: **shared documents + consistent prompt + individual evaluation + collaborative scoring** = better group decisions

GET STARTED TODAY

1

PICK YOUR NEXT RFP

Any upcoming evaluation
where you have 3+ responses

2

SET UP THE PROJECT

Upload RFP + scorecard.
Build the evaluation prompt
together with AI.

3

INVITE YOUR COMMITTEE

Share the project. Everyone
evaluates from the same
baseline.

The goal: **Set it up once. Evaluate better every time.**



QUESTIONS & DISCUSSION

Janette Roush · SVP, Innovation & Chief AI Officer

Skylar Clark · Senior Director, Partner Strategy

Agents of Change | Monthly Webinar Series | Brand USA