Challenge Set — Answer Keys and Facilitator Guidance

Full instructor guidance for every challenge in problem-statements.md: the model answer, the scoring guidance, the planted-defect lists, the role-play scripts, and — for every challenge — the Shipley discipline and the Dream DNA artifact it exercises.

> **How to use this file.** This is the facilitator's file, not the student's. It is separated from the problem statements on purpose: a challenge is destroyed the moment the learner reads its key. Run the cohort, collect the deliverables, *then* open the key for that number. The tier tells you the altitude of the expectation: an undergraduate deliverable is graded for method and completeness; a graduate deliverable for the quality of the defense; a professional-tier deliverable for whether a senior leader would act on it; a doctoral deliverable for the rigor of the research framing.
>
> **The DNA pointers.** Each entry ends with the Shipley discipline and the Dream DNA artifact the challenge exercises. DNA artifact paths are cited at `DreamLimited/dna` **v0.18.2** (the current pin; see `align/concept-to-system-map.md`). If a DNA version bump moves a path, update the pointer here — never the doctrine.

CH-01 — The Golden No-Bid

Discipline pointer. Shipley — bid/no-bid decision (the pursuit-decision gate). DNA — `rules/bid-no-bid.yaml`; `formulas/bid_no_bid_rubric.py`.

The arithmetic. Weighted pWin:

```
0.25×0.20 + 0.25×0.35 + 0.15×0.70 + 0.20×0.60 + 0.15×0.50
 = 0.050   + 0.0875  + 0.105   + 0.120   + 0.075
 = 0.4375
```

Composite = 0.4375 × \$1,800,000 = **\$787,500**. Both lines clear: pWin 0.4375 0.42, and composite \$787,500 0.42 × \$1,800,000 = \$756,000. The score *passes*.

The call. NO BID — or, at most, **BID WITH CONDITIONS** that include a real, named relationship-repair investment. The reasoning: the two weakest factors — customer relationship (0.20) and past performance (0.35) — are the two factors that are *hardest to fix in five weeks* and the two most load-bearing for an IDIQ where the customer is picking a partner it trusts. A passing composite built on the firm's two structural soft spots is a golden no-bid wearing a bid's clothes: the arithmetic is honest, and the judgment says the score is not *robust*. What would flip it: a specific, planned relationship move (a program-office briefing, a referenceable delivery win) with enough runway — not available in

five weeks. The final call belongs to the owner with standing; the process owner surfaces the soft-factor evidence so the decision is a check of boxes, not a mood.

Scoring guidance. A strong answer (1) shows the arithmetic; (2) makes a defensible call that is *not* "bid because it passes"; (3) names the two soft factors and why they matter more than the composite; (4) states the owner and the basis of the call. Award the point for a NO BID *or* a BID WITH CONDITIONS with a credible condition — the test is whether the student can explain why a passing number is not a passing decision. A "bid, it's above the line" answer fails the challenge even though the math is right.

CH-02 — Three Themes and a Discriminator

Discipline pointer. Shipley — win themes, discriminators, baseline features. DNA — `playbooks/03-proposal-strategy.md`; `prompts/personas/capture-manager.md`.

Model themes (each stated as the customer's outcome): 1. **"Earlier warnings."** BCRI's communities receive alerts hours sooner because the model fuses sensor and weather data in real time. (Written to Technical 40.) 2. **"Fewer false alarms."** Field crews chase real risks, not noise — the model's lower false-alarm rate protects scarce response capacity. (Written to Technical + the agency's published outcome.) 3. **"One shared platform."** Every BCRI office reads the same flood data from the same source — no silos, no reconciliation. (Written to the agency's stated "single shared data platform" outcome.)

Discriminators (claims Granite Peaks cannot credibly make, each with evidence): 1. The proprietary sensor-fusion model, with two published pilots — a specific, defensible claim Granite Peaks's "comfortable, slow-to-adopt" profile cannot match. 2. Native ingestion of existing flood-sensor networks, demonstrated in the pilots — a capability claim tied to evidence, not aspiration.

Baselines dressed as discriminators (the scan should catch these): "ISO-certified engineering process," "two decades of engineering services," "experienced project managers" — all claims any qualified offeror holds. They establish credibility; they do not move scores.

Strategy reading. The proposal must mirror the weights: Technical (40%) gets the best pages — the model and the platform theme. Management (30%) is the vulnerability: with no named program manager, the management theme is a promise without a person. The reading must flag that the PM must be named *before* proposal production, not during. Past Performance (20%) is carried by the two pilots; Price (10%) is a mid-range posture because the agency rewards innovation, not the cheapest bid.

Scoring guidance. Themes are graded on the *customer-outcome* test — a theme that reads "we offer X" is a feature, not a theme. Discriminators are

graded on the *could-a-competitor-copy-it* test. The strategy reading is graded on whether it mirrors the criteria in weight order and names the biggest threat (the unnamed PM).

CH-03 — The Ghost in the Room

Discipline pointer. Shipley — ghost themes; black-hat competitive reads. DNA — `rules/capture.yaml`; `prompts/critics/black-hat.md`.

Model ghost themes (Cardinal Point's strength stated so the customer draws the conclusion about Helixfield): 1. **The anti-silo ghost:** *"Our system is built to read the maintenance data FLC already collects — the data you already own, not a new silo."* Evidence: the pre-RFP technical exchange where the integration was verified; the program officer's industry-day complaint that the current system "doesn't talk to the maintenance data we already collect." The evaluator connects the dots to the incumbent's silo without the proposal saying it. 2. **The retention ghost:** *"The team we propose is the team that stays — the people who built the system run the system."* Evidence: two Helixfield engineers interviewing at FLC for direct-hire roles — a real signal that the incumbent's bench is turning over. Stated as a Cardinal Point commitment; the evaluator reads the incumbent's staffing risk.

The defensive read (what Helixfield runs at Cardinal Point): the incumbency baseline — *"six years of proven performance at FLC"* — which is a credibility claim, not a ghost; and a capacity ghost — *"can a small firm sustain a 24/7 pilot?"* Blunt it with: the pilot is bounded scope, the team is named, and the HUBZone/price-sharp posture is a strength, not a liability.

The drop rule. A ghost whose only evidence is rumor must be dropped — if the engineers-interviewing signal were hearsay with no captured event, the retention ghost is hollow posturing, and evaluators (who sit on the same discipline) read straight through it.

Scoring guidance. Ghost themes are graded on the *evidence* test — every theme must cite a specific captured engagement event or observed signal. A theme with no evidence fails regardless of how elegant it reads. The defensive read is graded on whether it names the competitor's likely move and the evidence that blunts it.

CH-04 — Reading the Field's Past Performance

Discipline pointer. Shipley — competitive analysis; incumbency assessment. DNA — `formulas/incumbency_pressure.py`; `rules/past-performance-retrieval.yaml`.

Model bidder profiles (strongest two): - **Granite Peaks Federal** — 14 awards, high domain match, no incumbency here. Likely win strategy: breadth + price. Soft spot: mixed record on *small* programs, and no DDWI-specific incumbency. Position: do not try to out-price them; beat them on domain-specific past performance and the small-business participation plan, which they will under-deliver. - **Helixfield Corporation** — the incumbent on the parent contract, 9 awards, 4 "satisfactory" ratings and one stale "exceptional." Likely win strategy: incumbency + continuity. Soft spot: a "satisfactory" streak signals comfort, not excellence; its past-performance edge is real but narrow.

Incumbency-pressure assessment. Helixfield's incumbency matters — DDWI explicitly values domain-specific past performance, and the incumbent owns the transition advantage. What neutralizes it: DDWI also scores a small-business participation plan, where Helixfield (large prime) is structurally weaker, and the "satisfactory" streak is beatable with a sharper, more domain-focused record. Incumbency is a factor to price and position against, not a wall to accept.

Recommendation on whom to position against. Position against **Helixfield** (the winnable incumbent) by out-scoring on domain-specific past performance and the participation plan. Do **not** try to out-price **Granite Peaks** — its price-sharpness and scale make a price war suicide.

Public-record boundary. Used: award data, CPAR-style ratings, the IDIQ awardee list, forecasts. Deliberately not used: anything non-public (rumor, inside contacts, competitor bid data). The boundary is absolute and is part of the grade.

Scoring guidance. Profiles are graded on whether the competitor's likely *win strategy* is inferred from the record, not just restated. The positioning recommendation is graded on the reasoning (why Helixfield and not Granite Peaks). The boundary note is graded on presence and correctness.

CH-05 — The Capture Plan That Has to Survive a Room

Discipline pointer. Shipley — capture plan; capture-readiness gate. DNA — `prompts/stages/capture.md`; `rules/capture.yaml`; `playbooks/03-proposal-strategy.md`.

Model capture plan (six parts): 1. **Win strategy** — themes (accessibility compliance, safe migration, a named PM), discriminators (Osprey's migration safety record; Atlas Run's integration bench), pricing posture (mid-range, best-value). 2. **Customer map** — the AVES program office decision-makers, the evaluators, and what each cares about (accessibility score, data-migration safety, the named PM). 3. **Solution** — a preliminary technical approach: current-state platform, migration path, accessibility remediation, with Osprey as the migration specialist. 4. **Team** — Atlas Run primes (integration + PM), Osprey fills

the migration gap; a teaming letter from Osprey is a condition of the readiness gate. 5. **Competitive landscape** — Granite Peaks and Helixfield likely; possibly a Blue Heron-led team; the price band and Helixfield's incumbency assessed. 6. **Engagement plan** — specific contacts by week before the RFP: an AVES industry day, a technical exchange with the migration program office, a reference call.

Readiness-gate criteria (the check that must pass before proposal work): (1) win strategy written and clear; (2) customer preferences understood and logged as engagement events; (3) teaming solidified (Osprey letter in hand); (4) key personnel named; (5) competitive landscape mapped with a price-band estimate; (6) preliminary solution aligned to the customer's mission. Outcome: proceed / defer / withdraw.

Scoring guidance. The plan is graded on completeness (all six parts present) and on whether the choices are *made* rather than deferred — the managing director's rule. The readiness-gate criteria must be written as pass/fail tests, not aspirations.

CH-06 — Defend This Gate

Discipline pointer. Shipley — capture-readiness gate (gate 4 of the BD ladder). DNA — `rules/shipley-capture-gates.yaml`; `formulas/readiness_score.py`.

Model recommendation. PROCEED WITH CONDITIONS. The gate check: pWin 0.55 0.42 (pass); composite $1,320,000 threshold (pass); **no disqualifying compliance gaps (fail — two open items)**. The conditions: (1) resolve the personnel substitution — confirm the proposed key person or update the Key Personnel table; (2) obtain the subcontractor letter, or drop the subcontractor claim from the proposal; (3) re-scope the capture lead's workload — an overstretched capture lead is the top delivery risk.

The three hardest questions the defense must anticipate: 1. *"Why spend another $60k when pWin is 0.55 and you have two open compliance items?"* — Answer: the pWin is defensible only because the two discriminators are real (name them, with evidence); the compliance items are the *reason* for the conditions, and the conditions are cheap to clear now and expensive to clear at the deadline. 2. *"What is the single strongest reason we win, and what proves it?"* — Answer: name the discriminator and its evidence; if the answer is a baseline, the gate fails. 3. *"What is the exit criterion — at what point in the six weeks do we stop?"* — Answer: a written stop-line (e.g., "if the compliance items are not resolved by week 3, or if a competitor's price band lands below our floor, we withdraw"). A gate without an exit is not a gate.

The live defense. The facilitator should push on the weak spots: the unknown price band, the overstretched capture lead, the two compliance items. A strong

team concedes the weakness, names the condition, and shows the evidence; a weak team becomes emotional — which is precisely the failure mode the gate exists to catch.

Scoring guidance. The deck is graded on the four gate parts being explicit (trigger, criteria, evidence, owner). The defense is graded on whether it can answer a fact — a team that cannot name its discriminator or its exit criterion fails regardless of presentation quality.

CH-07 — The Pink Team Finds What the Writers Missed

Discipline pointer. Shipley — color teams (pink, red). DNA — `rules/color-team.yaml`; `prompts/critics/pink-team.md`; `prompts/critics/red-team.md`.

The planted-defect list (for the instructor)

The draft excerpt handed out with the challenge contains **eight planted defects**. Five are Pink findings (compliance/coverage); two are Red findings (persuasiveness/score); one is both. The students are not told which are planted — finding all eight is the pass bar.

1. **Page-count violation (Pink).** The Technical volume draft runs **54 pages** against the 50-page Section L limit. A compliance disqualifier.
2. **Missing agency-specific certification (Pink).** The NBHA data-use certification — the third required certification — is absent from the representations section. Easy for an evaluator to find; fatal.
3. **Key-personnel substitution in prose (Pink/Red).** The management section names **Jordan Reyes** as the proposed program manager, but the Key Personnel table names **Sam Okonkwo**. The substitution is made in prose with no table update — a compliance break *and* a credibility break.
4. **Ghost-theme rule violation (Red).** The technical approach explicitly names a competitor: *"unlike Granite Peaks Federal's aging platform."* Naming the competitor turns a ghost theme into an attack and violates the discipline.
5. **Baseline dressed as a discriminator (Red).** The proposal calls *"our ISO 9001 certification"* a key discriminator. ISO 9001 is a baseline any qualified offeror holds.
6. **A Section L "shall" with no home (Pink).** Section L requires a **Transition-In Plan**; the draft has no transition section anywhere in the outline.
7. **Price inconsistency (Pink/Red).** The Technical volume references **$1,900,000**; the Cost volume shows **$2,150,000**. A cost–technical reconciliation failure.
8. **Unverified SAM claim with no owner (Pink).** The draft asserts *"SAM registration active"* with no evidence, and the compliance-matrix row has **no owner**.

Model Pink report

Every finding is written as *factor + evidence + required revision.* Examples: finding 1 → "Compliance / Section L page limit — Technical volume is 54 pages vs. 50 required; reduce to 50 before Red." Finding 4 → "Technical / Section M persuasiveness — the ghost theme names Granite Peaks; rewrite as a ghost ('a platform that is not tied to legacy systems') so the evaluator draws the conclusion."

Model Red scores

- **Technical factor: Yellow (Marginal).** Significant weaknesses: page overrun, a ghost-theme violation, a baseline dressed as a discriminator, and a price inconsistency. Must be resolved.
- **Management factor: Yellow (Marginal).** Key-personnel substitution and the Transition-In gap. Must be resolved.

Disposition. NOT submittable as-is. The page-limit violation (finding 1) and the missing certification (finding 2) are compliance disqualifiers; the price inconsistency (finding 7) blocks the cost–technical reconciliation the Red Team requires. Findings 1–3 and 6–8 must be resolved before Red; findings 4–5 are Red-level revisions before Gold.

Scoring guidance. The pass bar is finding **all eight** planted defects as specific findings. Partial credit: a student who finds the page overrun and the missing certification but misses the ghost-theme violation and the baseline-dressed-as-discriminator has done a compliance pass but not a strategy pass. Vague findings ("technical is weak") are rejected outright per the discipline.

CH-08 — The RFP That Was Written to Be Misread

Discipline pointer. Shipley — compliance matrix; RFP decode. DNA — `playbooks/compliance-matrix/compliance-matrix.yaml`; `rules/section-l-format-compliance.yaml`; `formulas/compliance_parser.py`.

The planted-traps list and resolutions

The RFP summary contains **eight traps**. The resolution column is the professional behavior:

1. **50 vs. 75 page conflict.** Section L (instructions) says 50; Section M (evaluation) references a 75-page volume. **Resolution:** Section L is the instructions and governs format; comply with **50 pages**. Flag the ambiguity in a question.
2. **Portal contradiction.** Section L names the OARP portal; **Amendment 2 changes the portal. Resolution:** the amendment supersedes

— submit to the amended portal. The trap kills teams that stop reading at Amendment 1.

3. **Past-performance count: 4 vs. 3 references.** Section L says max 4; Section C says max 3. **Resolution:** when instructions conflict with the statement of work, the **more restrictive** rule governs — submit **3**; confirm via the questions deadline.
4. **11-pt template vs. 12-pt font rule. Resolution:** the format rule governs; set the template to 12-pt before writing. Do not submit an 11-pt document because the template shipped that way.
5. **Two-deadline trap.** Body text says proposals due 2026-09-24; the **portal says 2026-09-29**. **Resolution:** the portal is authoritative for submission; treat 09-24 as the questions/body deadline and confirm the true submission deadline from the portal and amendments. Set the internal deadline ~1 week before the portal date.
6. **Section J data-security form not referenced elsewhere. Resolution:** still required — add it to the matrix and the package; a form that appears nowhere else is the easiest one to miss.
7. **Missing certification in Amendment 2. Resolution:** complete the amended certification; verify the form's revision date against the amendment.
8. **SAM active requirement. Resolution:** verify the registration is active and current *at time of offer* — an inactive record is a disqualifier regardless of content.

Model compliance matrix

Eight-plus rows, each a "shall" with source (the exact section), satisfaction point (where the response satisfies it), and an owner. The traps themselves become rows (e.g., row: "Amended submission portal — Amendment 2 — submit via the amended portal — capture lead").

Model deadline plan

- **Questions deadline:** 2026-09-10 (submit written questions before this date).
- **Submission deadline:** 2026-09-29 (portal — authoritative).
- **Internal deadline:** 2026-09-22 — a full week before the portal close, leaving room for review, fixes, and the mechanical act of submitting.

Scoring guidance. The pass bar is catching **all eight traps**. The matrix is graded on completeness and on the *source* column being exact. A matrix that treats the 50-page limit and the 75-page reference as both-satisfiable fails the audit — the resolution matters as much as the row.

CH-09 — $240,000 to Win On

Discipline pointer. Shipley — price-to-win. DNA — `formulas/pricing_to_win.py`; `rules/pricing-floor.yaml`; `playbooks/pricing/cost-volume.yaml`.

The cost buildup. Senior: 900 hrs × $118 = **$106,200**. Junior: 450 hrs × $82 = **$36,900**. Labor = **$143,100**. Indirect (32%) = **$45,792**. Fully loaded cost = **$188,892** the $189,000 floor. Fee (8%) = **$15,111**. Price at cost-plus-fee = **$204,003**.

The price-to-win position. The floor is ~$189k; the ceiling is $240k; the estimated competitive range is $215k–$245k; Helixfield is well-liked and expected near the ceiling; evaluation is best-value with Price weighted 30%. The defensible recommendation is **$215k–$220k** — the **low end of the range**, above the floor with real margin (~$26k–$31k, 14–16%), carrying a strong technical story to cover the past-performance gap. Do **not** bid at the floor: in a best-value competition, a floor bid signals desperation and forfeits the margin the firm needs. Do **not** chase the ceiling: the incumbent owns the high ground. The story: *"We price at the low end of the competitive range because our lean structure lets us — and we spend the saved margin where it matters, on the technical approach."*

Sensitivity. The line you will not cross: **below cost** — at $189k you win a contract that bleeds you. Below ~$195k the fee collapses under ~3% and the risk-adjusted value turns negative — reconsider. A sharp price of ~$204k (cost + 8% fee) is a legitimate *alternative* play that undercuts the field while holding fee — but in best-value against a well-liked incumbent, the ~$215k position with a technical story is the stronger defense.

Scoring guidance. The build is graded for arithmetic correctness and for the floor being identified as the *no-cross* line. The position is graded on whether it sits inside the range, above the floor, and is justified as a position with the story — not on whether it matches the model exactly. A floor bid or a ceiling bid without justification fails the price-to-win test.

CH-10 — Two Sides of One Table

Discipline pointer. Shipley — negotiation. DNA — `prompts/stages/review.md`; `rules/pricing-guardrails.yaml`.

The role-play script (for the instructor)

Side A — Meridian Anchor (the student team). - **Real interest:** keep the same program manager on the work, and protect the optional period's profitability. - **Opening position:** hold the $148 blended rate. - **Concessions, in order:** (1) move to $146 on a 2-year commitment; (2) move to $145 with a firm-fixed rate for the full option period and a cap on ODCs; (3) accept $144 if

BCRI guarantees the program manager in writing. - **Reservation price:** $143 — below this, the optional period is unprofitable. - **The concession you will not make:** the program manager. That is the leverage BCRI is really paying for. - **BATNA:** let the option lapse and bid the follow-on fresh (costly for both sides) or walk to a shorter bridge agreement at the current rate.

Side B — BCRI (the facilitator). - **Real interest:** budget certainty for the option period, and keeping the same program manager — the agency keeps asking for them. - **Opening position:** demand $139 (−6%). - **Concessions, in order:** (1) accept $145 (−2%) on a firm-fixed rate for the full option; (2) accept $144 with a cap on ODCs and a multi-year pricing commitment; (3) accept the current $148 if Meridian guarantees the PM and adds a small scope guardrail. - **Walk-away:** exercise the option at the current rate for one more year, then recompete — which BCRI would rather avoid because recompete is expensive and risks losing the PM. - **The concession BCRI will not make:** letting the PM leave the contract.

Model prep sheet

Interests (not positions), BATNA (Meridian: walk to the follow-on; BCRI: recompete), ZOPA (roughly $143–$148, with a sustainable deal likely $144–$146), reservation price ($143), and the three concessions before the one that is off the table (the PM). The post-negotiation memo grades the *relationship* outcome as much as the rate: a deal at $145 with the PM guaranteed and a firm-fixed rate is a win for both sides; a deal at $139 that loses the PM or the profitability is a loss dressed as a win.

Scoring guidance. The prep sheet is graded for the vocabulary (interests vs. positions, BATNA, ZOPA, reservation price). The live negotiation is graded on preparation beating improvisation: a team that states a position and a reason, concedes in order, and protects its real interest scores higher than a team that "wins" the rate and loses the relationship.

CH-11 — The Oral That Almost Won

Discipline pointer. Shipley — orals; discussions. DNA — `prompts/stages/review.md`; `playbooks/03-proposal-strategy.md`.

The model critique findings

The transcript contains four broken rules. Each finding is *rule + evidence + correction*:

1. **No improvisation of strategy on stage.** The technical lead invents a "new 10% discount" in the opening. A price commitment not in the written proposal is a hole in the story — and a price change mid-oral is

a strategy change mid-competition. **Correction:** price questions route to the pricing lead; all presenters are briefed on the exact written price position before the room.

2. **Bring the people you proposed.** The program manager on stage is not the named key person. Substituting key personnel at the oral is a credibility hit evaluators remember. **Correction:** the named PM presents, or (with agency consent) the written proposal is amended before the oral.
3. **Roles and slides are pre-decided.** The pricing lead spends six minutes on a slide that is not in the deck, blowing the time budget. **Correction:** freeze the deck; rehearse to a stopwatch; an external observer calls time.
4. **The oral is the proposal in person.** The closing theme contradicts the written executive summary. **Correction:** reconcile the oral script to the written themes before the next rehearsal.

Model rehearsal plan. Roles pre-decided (who speaks, who answers, who observes); a stopwatch drill run twice; an external observer playing a hostile evaluator; and a **Q&A bank** built from the evaluation factors, the risk register, the price position, the past-performance citations, and the key personnel — each question with a primary and a backup responder.

Scoring guidance. The critique is graded on specificity — a finding that names the transcript moment and the rule it broke earns full credit; "the oral was unfocused" earns none. The rehearsal plan is graded on whether it includes the hostile-observer drill and the Q&A bank, which are the two moves that actually improve an oral.

CH-12 — Rebuild the Loss From the Debrief

Discipline pointer. Shipley — debrief; lessons learned. DNA — `rules/past-performance-retrieval.yaml`; `playbooks/06-post-submission.md`.

Model reconstruction

- **What landed:** Management (4/5) — the named PM and the org story worked. Price was competitive (4/5).
- **What did not:** Past performance (2/5) — the debrief's blunt line: *"your references did not match the work you proposed."* The references were adjacent, not relevant. Technical (3/5) was solid but not differentiating — the discriminators did not move the panel the way the win strategy hoped.
- **The synthesis:** the win strategy leaned on team and technical; the actual evaluation weighted past performance more heavily than the model did, and the reference mismatch made the firm's one structural weakness the deciding factor. A theme cannot compensate for evidence that does not match the work.

Updated pWin factors

Original factor scores (consistent with the 0.62 bid/no-bid): topic fit 0.80 × 0.25; team 0.75 × 0.20; past performance **0.55** × 0.20; price 0.50 × 0.20; capacity 0.40 × 0.15 → **pWin 0.62**. Re-score past performance to **0.30** (references did not match): new pWin = **0.57**. The lesson is not "we were overconfident by 0.05" — it is that the past-performance *reference-matching* is a capture deliverable, not a proposal afterthought, and the factor model under-weighted it relative to how this agency scores.

Study design (doctoral)

Research question: *Does past-performance-reference matching moderate the predictive validity of the factor-weighted pWin composite against actual award outcomes?* Variables: pWin factor scores (predictor), award outcome (criterion), reference-match quality coded from debriefs (moderator). Validity threats to name: only losing offers receive detailed debriefs (survivorship), self-reported scoring, small N, agency heterogeneity. A series of such reconstructions is the raw material of the ORBITAL outcome-data research agenda.

Scoring guidance. The reconstruction is graded on whether it maps debrief findings to the win strategy and the compliance matrix, not just restates the scores. The factor update must show the arithmetic. The doctoral study design is graded on naming a research question, variables, and at least one validity threat.

CH-13 — The Cost Volume the Auditor Will Read

Discipline pointer. Shipley — cost/price volume. DNA — `rules/pricing-guardrails.yaml`; `formulas/loe_burdened_rate.py`. (FAR cost principles at GC 540 altitude; the FAR is the law, not the method.)

The audit findings (model memo)

The drafted cost volume contains five problems; the memo must identify each and the rule it violates:

1. **$40,000 "executive incentive tied to award" — unallowable.** Compensation contingent on the award is not an allowable cost under the FAR cost principles (FAR 31.205-6). Remove it entirely.
2. **$12,000 industry-conference trip with no agenda — not allocable.** Travel costs must be reasonable and allocable to the contract; a trip with no agenda has no demonstrated benefit to the work (FAR 31.205-46). Reclassify with a written agenda and a stated tie to the contract, or disallow.

3. **$6,000 "employee entertainment" — unallowable.** Entertainment costs are expressly unallowable (FAR 31.205-14). Remove.
4. **A rate that does not match the disclosed accounting practice — a CAS/auditability problem.** Charging an indirect rate inconsistent with the firm's disclosed practice is a cost-accounting-practice change without approval (CAS 48 CFR 9904) and makes the entire volume unauditable. Rebuild the volume on the audited 32% / G&A 12% structure.
5. **A one-sentence basis of estimate — unauditable.** DCAA expects a basis of estimate that shows the hours, the labor categories, and the rationale — the *what, how, and why* of the estimate. A single sentence per line is not an estimate; it is a hope.

Model basis-of-estimate

For the two largest labor lines: the number of hours per task (decomposed from the work plan), the labor category and rate (matching the disclosed structure), the specific task the hours support, and the source of the estimate (prior similar work, historical actuals, or engineering judgment stated as such).

Scoring guidance. The memo is graded on naming the *rule* behind each correction, not just "this is wrong." The corrected volume is graded on internal consistency — one indirect structure applied throughout. The B.O.E. is graded on whether it would survive an auditor asking "where do these hours come from?"

CH-14 — One Deal, Two Companies, Seven Axes

Discipline pointer. Shipley — teaming, JV, subcontracting. DNA — `rules/sba-size-standards.yaml`; `rules/naics-alignment.yaml`; `rules/capture.yaml`. (No dedicated teaming artifact in DNA at v0.18.2 — the closest is the size/affiliation rule; the teaming canon stays at concept altitude.)

The role-play script (for the instructor)

Side A — Blue Heron Services (the student team). - Real interests: prime the contract (the set-aside requires it), meet the 40% statutory workshare floor, and capture the fee split that recognizes that its eligibility is what makes the pursuit possible at all. - **Opening position:** prime; 45/55 workshare (45% Blue Heron); fee split 60/40 in its favor. - **Concessions, in order:** (1) workshare down to the 40% floor in exchange for the fee split; (2) fee split to 55/45 if Northwind guarantees the technical approach stays on its patented methods; (3) a reuse license to Northwind's methods for other federal work if Northwind grants a license-back for Blue Heron's domain deliveries. - **The concession it will not make:** losing the prime. Without primacy, the set-aside is lost. - **BATNA:** team with a different technical partner; or wait for a

set-aside it can staff more of.

Side B — Northwind Systems (the facilitator). - **Real interests:** keep the 14-person bench fully loaded, control the technical approach and the IP on its patented methods, and get a workshare that justifies bringing the full bench. - **Opening position:** workshare 50/50; fee split 50/50; strong IP protection on the patented methods; license-back terms narrow. - **Concessions, in order:** (1) accept Blue Heron as prime (it has no eligibility to contest this); (2) workshare to 45/55 (Blue Heron 45%) if its people stay loaded; (3) fee split to 55/45 in Blue Heron's favor if the technical approach remains Northwind-controlled. - **The concession it will not make:** title to its patented methods. It will grant a license, not a transfer. - **BATNA:** find a different set-aside-eligible partner, or prime an unrestricted competition it can win without a partner.

Model term sheet and ORBITAL

Term sheet: prime = Blue Heron; workshare 40% Blue Heron (statutory floor) with a defensible split (45/55); fee split reflecting that eligibility is a load-bearing contribution (55/45 or 60/40); IP — Northwind retains title to its methods, grants Blue Heron a license for the contract and a limited license-back; key personnel named; an exit provision for material breach. ORBITAL **Resources** axis: the two firms, named leads, and the eligibility each brings. **Budget** axis: the $6.5M value, the workshare cost split, the fee split, and the pursuit cost.

Scoring guidance. The term sheet is graded on the four pressure points being addressed (prime, workshare, fee, IP). The outcome memo is graded on whether it records what each side *got* and *conceded*, and whether the deal is built to last the full three years — a deal that trades the prime for fee is a loss even if the arithmetic favors it. The ORBITAL is graded on the two axes holding the partnership's structure and money.

CH-15 — From Seven Axes to Four Volumes

Discipline pointer. Shipley — proposal structure off the capture position. DNA — `rules/orbital-to-proposal-map.yaml`; `playbooks/volumes/technical-volume.yaml`; `playbooks/volumes/management-volume.yaml`.

Model axis → section mapping

ORBITAL axis	Feeds
Objective	Executive summary; Technical volume opening (the problem and the success definition)

ORBITAL axis	Feeds
Resources	Management volume (team, key personnel, the STTR effort split); Technical volume (staffing); Past Performance volume (the people's relevant record)
Budget	Pricing volume (cost/budget + justification); Management volume (indirect structure, fee)
Indicators	Management volume (quality assurance / performance measurement)
Transport	Technical volume (delivery approach — prototype demo, reporting); Management volume (delivery channel)
Activities	Technical volume (work plan, schedule); Management volume (project plan)
Logistics	Technical volume (facilities, compliance); Management volume (assumptions and constraints)

Model gap list

The four volumes require things the ORBITAL does not hold. The gaps, and where they would surface: 1. **Past-performance evidence.** The ORBITAL names people but not their relevant prior contracts; the Past Performance volume needs citations. Gap surfaces in the Pink Team review. 2. **Commercialization / Phase II thesis.** The ORBITAL's Objective mentions "a credible Phase II thesis" but holds no content; the Commercialization Plan needs it. Gap surfaces in Technical/Commercialization drafting. 3. **Basis of estimate detail.** The Budget axis gives totals, not the hours-and-rates build-up the Pricing volume requires. Gap surfaces in the cost volume. 4. **Key-personnel bios and effort split.** The Resources axis names roles but not the 40/30 STTR split as a narrative; the Management volume requires it stated.

Model volume outline

Technical (15 pages) leads with the objective and the model (the highest-weighted criterion); Management names the PI, the effort split, and the QASP (Indicators); Past Performance maps each citation to the work; Pricing shows the budget aligned to the Activities axis — budget and work tell one story.

Scoring guidance. The mapping is graded on coverage (every axis lands somewhere) and on the gap list being specific — naming the volume and the

review where the gap would surface. The outline is graded on whether it mirrors the evaluation criteria in weight order.

CH-16 — The Portfolio and the Line

Discipline pointer. Shipley — pWin; scoring. DNA — `formulas/pwin_scorer.py`; `formulas/pwin-shipley.formula.yaml`; `formulas/risk_score.py`.

The arithmetic

Opportunity	Composite	Threshold (0.42 × value)	Call
A — STTR Phase I	0.77 × \$250k = **\$192,500**	\$105,000	**BID**
B — HUBZone set-aside	0.58 × \$180k = **\$104,400**	\$75,600	**BID** (park below)
C — Large IDIQ	0.30 × \$2.4M = **\$720,000**	\$1,008,000	**NO BID**
D — Data-hosting renewal	0.82 × \$120k = **\$98,400**	\$50,400	**BID** (light touch)
E — New-market entry	0.18 × \$900k = **\$162,000**	\$378,000	**NO BID**

Note: opportunity **C** fails the pWin line ($0.30 < 0.42$) *and* the composite line (\$720k < \$1,008k). Opportunity **E** fails both lines too.

The hard calls

- **C — the trap of the big number.** The \$2.4M value makes a \$720k composite look meaningful, but the probability line is the point of the threshold: an unstaffable, no-relationship pursuit at 0.30 fails both lines. The composite "looks" big because the value is big; the threshold exists to catch exactly this. **NO BID**, in writing.
- **E — the tempting composite.** \$162k composite is larger than D's \$98k, but E fails both lines and the relationship factor is zero. A composite is not a recommendation; the threshold and the judgment are. **NO BID.**

The portfolio line

Given Ravonics's capacity of **one active pursuit**: advance **A** (best fit, pWin 0.77, the one active pursuit), take **D** as the light-touch renewal (light capacity cost), park **B** (scoreable, eligible, but lower value — defer to the next slot), and

decline **C** and **E** in writing. The portfolio is a deliberate allocation of scarce attention, not a ranking of composite sizes.

Scoring guidance. The sheet is graded on arithmetic correctness and the gate calls following from the stated threshold. The hard-call paragraphs are graded on whether the student can explain *why* a big composite is still a no-bid — the threshold discipline, not the number, is the test.

CH-17 — The Final Proposal Revision

Discipline pointer. Shipley — FPR / discussions. DNA — `playbooks/03-proposal-strategy.md`; `prompts/stages/review.md`.

Model FPR strategy memo

- **Weakness 1 (no named PM) — FIX, highest priority.** Name a single accountable program manager in the management section with a one-line bio. This is the cheapest, highest-leverage score move in the entire FPR: an evaluator who reads "one person owns this" resolves the management weakness immediately.
- **Weakness 2 (marginally relevant references) — FIX with a relevance bridge.** Replace the least-relevant reference with the closest match, and add one paragraph per reference mapping it to the statement of work. The debrief lesson from CH-12 applies: reference *matching* is a capture deliverable.
- **Weakness 3 (price above range) — SHARPEN.** Move from \$1.15M to **\$1.08M–\$1.10M** — still above the \$1.02M floor, with a one-page value story (lean cost structure, the technical approach, the HUBZone/price-sharp posture). The line you will not cross: **\$1.02M**.
- **What stays.** The technical approach (no feedback), the win themes, the discriminators. An FPR is a compressed cycle, not a rewrite — changing what the agency did not criticize invites new risk.

Model review plan (12 days)

Day 1–2: strategy + the PM fix. Day 3–6: write the 10 pages. Day 7–8: **Pink pass** — compliance against the FPR instructions (page limit, format, forms). Day 9: **Red pass** — does the revision answer all three weaknesses? Day 10: **Gold** — the executive submit call. Day 11: white-glove + final verification. Day 12: submit **early**, with the upload verified, not assumed.

Scoring guidance. The memo is graded on the *proportion* discipline — fix what the agency criticized, protect what it did not. A plan that rewrites the whole proposal fails the FPR test. The review plan is graded on whether the color sequence runs compressed but intact, never skipped.

CH-18 — Rescue the Derailed Pursuit

Discipline pointer. Shipley — capture gates; organizational governance. DNA — `rules/capture.yaml`; `formulas/capture_readiness_index.py`.

Model root-cause analysis

The derailment is a catalog of skipped gates, each with a specific mechanism: 1. **No bid/no-bid.** The pursuit advanced on a "hallway yes" — the qualify gate was never run, so the pursuit was never honestly scored. This is the original sin. 2. **No scoring.** Without a score, no threshold conversation happened — resources were committed before the arithmetic. 3. **Half-built compliance matrix.** The "shall" audit never ran to completion; two requirements are confirmed missing. The matrix is the backbone of every later review; a half-built matrix means every later review was built on sand. 4. **No approved outline.** Writers drafted against an outline nobody approved — production ran ahead of governance. 5. **Missed internal deadlines.** No internal-deadline rule existed, so the invisible third deadline was never set. 6. **Single-point-of-failure.** The capture lead left and took the plan (such as it was) with them — the pursuit lived in one person's head, the exact failure the ORBITAL is designed to prevent.

Model recovery decision

Recovery is a **gate decision, not a rescue mission.** Criteria: (1) if a score run *now* clears the threshold, and (2) the two missing "shall" items are fixable within the $12k budget, then **RESTART** as a disciplined 4-week compressed cycle; otherwise **WITHDRAW** and record an honest no-bid. The re-entry gate: score it now; rebuild the compliance matrix first (it is the backbone); approve the outline at a review; name one accountable capture manager; and cut scope to a compliant, competitive minimum — the proposal is no longer the original ambition, it is a disciplined version of it that can be delivered.

Research framing (doctoral)

The case as a data point: variables to code — gate compliance (was each gate actually run?), documentation quality (did the plan live in a structure or a person?), single-point-of-failure (was there a named owner + a written plan?). Counterfactual: *would a scored bid/no-bid have stopped this pursuit at week zero?* A series of such cases would test whether derailments share a signature — and whether the gate discipline, not the proposal skill, is the variable that predicts survival.

Scoring guidance. The root-cause analysis is graded on mapping each failure to a specific gate or structure, not on generic "poor planning." The recovery

is graded on being a *decision with criteria* rather than a sprint. The doctoral framing is graded on the counterfactual and the variables.

CH-19 — The Portal Closes at Five

Discipline pointer. Shipley — submission mechanics. DNA — `rules/submission.yaml`.

Model incident log

Time	Action
4:10 PM	Upload begins.
4:18 PM	Validation error: one attachment 12 MB over the 50 MB limit; the signed form is missing.
4:20 PM	Capture lead calls the contracting officer — confirms the portal closes mechanically at 5:00; no grace.
4:25 PM	Team splits: admin recompresses the oversized attachment; writer retrieves the *amended* form (Amendment 3) and has it signed; capture lead re-verifies the package against the amended requirements.
4:55 PM	Package re-uploaded; validation passes.
4:59 PM	Submission confirmed; confirmation email saved.

The professional decisions in the log: never submit a non-compliant package "to get something in"; never chase polish — a compliant, slightly-less-polished submission beats a late or non-compliant one; and the re-verification *after* Amendment 3 is the move that saved the day.

Model post-mortem

Root cause: the internal deadline was set for "the day before," but nobody **re-verified the package after Amendment 3** — the amendment changed the required form's revision date, and the team kept the old form. The three fixes: (1) a **delta check** after every amendment (re-verify the package against the new requirement); (2) an internal deadline that includes a **final mechanical upload rehearsal** the day before — the same "third deadline" the doctrine teaches; (3) a **pre-submission checklist** where every form is checked against its current revision date.

The invisible third deadline

The internal deadline exists to absorb exactly this day — a validation error, a changed form, a portal jam. The pursuit is not ready because the calendar says so; it is ready because the review gates say so.

Scoring guidance. The log is graded on the decisions being professional (compliant over polished, verified over assumed). The post-mortem is graded on naming the amendment as the root cause and the delta check as the fix — the specific lesson, not a generic "communicate better."

CH-20 — Four Opportunities, One Team

Discipline pointer. Shipley — market segmentation; pursuit decision; portfolio. DNA — `rules/bid-no-bid.yaml`; `formulas/risk_score.py`.

Model portfolio recommendation

- **Advance: A (STTR Phase I)** — best fit (5/5), pWin 0.77, the firm's one active pursuit. This is the bet that builds the cosmology loop: a won-and-delivered Phase I becomes the past performance that raises every later pWin.
- **Take light-touch: D (data-hosting renewal)** — pWin 0.82, $3k capture cost, light capacity. A low-effort, high-probability renewal that keeps the revenue base turning while A is in play.
- **Park: B (HUBZone set-aside)** — scoreable (0.58, composite $104k) but lower value and full capacity cost. Defer to the next slot; park, do not decline.
- **Decline in writing: C (large IDIQ)** — pWin 0.30 fails the threshold, the capture cost is high ($25k), and it would require hires the firm cannot make now. The large value is the trap.
- **Decline in writing: E** — no relationship, pWin 0.18; a disciplined "no."

Model no-bid defense (for C, the attractive one)

"The $2.4M value is real, but the pursuit fails the threshold at 0.30, requires hires we cannot make in the window, and would consume the entire capture budget against a 30% probability. Chasing the biggest number on the board is exactly the failure the threshold exists to prevent. NO BID — and we will re-sense the IDIQ when we have the bench to staff it."

Scoring guidance. The recommendation is graded on the gate reasoning at each rung (eligibility → score → capacity) and on the written no-bid being a *defense*, not a dismissal. A student who ranks by composite size alone fails the portfolio test — the capacity constraint is the point.

CH-21 — The Customer Who Has Never Heard of You

Discipline pointer. Shipley — customer engagement; capture. DNA — `prompts/stages/discovery.md`; `prompts/personas/capture-manager.md`.

Model engagement plan (DREAM-framed)

- **Discover (months 1–2).** Research OARP's published priorities and its data gaps; map the program office and the evaluators. **Intelligence objective:** learn which gap the program staff care about most — the raw material for a ghost theme.
- **Relate (month 3).** Attend the industry day; request a one-on-one technical exchange; follow up with a thank-you and a capability summary. **Intelligence objective:** become a known name; collect the signals — what OARP says it cannot get from its current partners.
- **Empower (months 4–5).** Share the published paper; offer a 30-minute capability briefing on the modeling method. **Intelligence objective:** demonstrate credibility and collect the *"we wish someone would…"* statements that become ghost-theme evidence.
- **Advise (month 5).** Provide a small, no-strings analytical insight — a free look at one named data gap. **Intelligence objective:** become a source the program office turns to; convert a contact into a referenceable engagement.
- **Motivate (month 6, pre-RFP).** Position as the partner of choice; confirm the relationship; lock the collected intelligence into the win strategy before the RFP drops.

The relationship-risk note

If OARP engages Granite Peaks and not Cardinal Point: the fallback is (1) a teaming angle — offer the modeling capability to the likely winner, or (2) an honest downgrade — lower the pursuit's priority or no-bid rather than chase a relationship the firm never built. Never fabricate engagement; a pursuit that is not genuinely engaged decays.

Scoring guidance. The plan is graded on specific, named contacts per month (not "network more") and on each engagement carrying an *intelligence objective*. The relationship-risk note is graded on naming a real fallback — the discipline is honest about the decay.

CH-22 — The Executive Summary That Must Carry the Ship

Discipline pointer. Shipley — executive summary; proposal strategy. DNA — `playbooks/03-proposal-strategy.md`; `prompts/personas/capture-manager.md`.

Model executive summary (structure)

One page, mirroring the criteria in weight order: 1. **Open with the mission and the one-line win** — the outcome BCRI gets, in its words, not the firm's capabilities. 2. **Technical (40%) first and fullest** — the sensor-fusion model, the shared platform, and the discriminator: two published pilots. This is the strongest section because it is the heaviest-weighted criterion. 3. **Management (30%)** — the named program manager and the accountability story; the fix from CH-02 (a named PM) must appear here. 4. **Past Performance (20%)** — the two pilots, with a one-line relevance bridge; the weakness addressed, not hidden. 5. **Price (10%)** — the mid-range posture, framed as best-value strength, not a discount. 6. **Close with the discriminator repeated** and a commitment sentence.

The cut-and-lead note

What was cut: team bios, facilities description, detailed pricing tables — they live in the volumes. What leads: the technical outcome, because Technical carries 40% of the score and the evaluator's mental model is built in the first paragraph. The past-performance weakness is handled, not hidden, because a one-page summary that hides the weakest criterion forfeits the chance to frame it.

Compliance check

Page count verified against the agency template (one page); font set to 12-pt; the template's margins and headers honored. Compliance before style — the discipline of Section L applies to the summary too.

Scoring guidance. The summary is graded on criteria order, theme threading, and the page limit being a hard rule. A summary that reads like a capability list rather than a customer-outcome story fails the win-theme test even if it fits on the page. The cut-note is graded on whether the student can say what was sacrificed and why.

CH-23 — The Agent's 0.62

Discipline pointer. Shipley — bid/no-bid under AI-assisted capture. DNA — `rules/bid-no-bid.yaml`; `formulas/pwin_scorer.py`; `rules/past-performance-retrieval.yaml`. Module — `modules/human-ai-collaboration.md`; the **override log** is the wave's core artifact.

The planted-error list (for the instructor)

1. **Hallucinated past-performance citation.** The AI's memo cites "a 2024 OARP-funded sea-ice forecasting study delivered by Cardinal Point."

No such contract exists in the firm's record (the real record is one domain-adjacent state-agency analytics delivery; nothing appears in the published award record). The agent either fabricated the reference or misattributed a partner's work. This inflates the past-performance factor from the real **0.25** to the claimed **0.60** (+0.07 to the composite pWin).

2. **Skipped cost-share compliance.** The NOFO requires a **mandatory 25% non-federal cost share**; the AI's plan states "no cost share required." Cardinal Point has no committed match funding. A bid without the match commitment is **non-compliant** and is rejected without evaluation — a gate-stopper, not a risk factor.

Model answer

- **Override 1 (correction).** Remove the fabricated past-performance citation; re-score past performance **0.60 → 0.25**. Evidence: the real record in the case file; the citation does not exist in the published record. This is a correction, not a judgment call — the agent was factually wrong.
- **Override 2 (correction).** Add the cost-share requirement to the compliance matrix as a hard "shall"; the mandatory 25% match is not optional. Optionally re-score price/cost **0.40 → 0.35** for the burden of raising the $100k match, or carry it purely as a condition. Either is defensible; the compliance miss is the finding either way.
- **The re-scored arithmetic** (with the price/cost re-score):

  ```
  0.25×0.80 + 0.20×0.70 + 0.20×0.25 + 0.20×0.35 + 0.15×0.50
   = 0.200   + 0.140   + 0.050   + 0.070   + 0.075
   = 0.535  → pWin ≈ 0.54; composite ≈ $214,000 ≥ 0.42×$400k = $168,000
  ```

  The number still clears the threshold — which is precisely the trap. The corrected composite *passes* while the pursuit is **not bid-ready** because the compliance gate is not met.
- **The call. BID WITH CONDITIONS** — (1) secure a written non-federal cost-share commitment of 25% ($100,000) before the readiness gate, or withdraw; (2) rewrite the past-performance section from the real record (the one adjacent delivery plus any verifiable research-partner relationship); (3) re-run the full compliance matrix against the NOFO's actual terms. A NO BID is also defensible if the match funding cannot plausibly be raised in the window — the conditions are the test.
- **Division of labor.** The agent computed the arithmetic from the inputs it was given; the human owns the **facts** (what is actually in the firm's record), the **compliance requirements** (what the NOFO actually says), and the **final gate call**. The 0.62 was a hallucination wearing a score's clothes: the arithmetic was honest, the inputs were not.

Scoring guidance. Full credit requires (1) both planted errors found and

classified as *corrections*; (2) the re-score shown as arithmetic with the real past-performance factor; (3) a BID WITH CONDITIONS call (or NO BID with the cost-share stated as the disqualifier) with the conditions explicit; (4) an override log that separates machine-computed inputs from human-owned facts. A student who accepts the 0.62 and bids fails the AI-accountability test even though the composite clears — finding the fabrication is the point.

CH-24 — The Ghost Pass That Missed the Collision

Discipline pointer. Shipley — color teams (red); ghost themes. DNA — `rules/color-team.yaml`; `prompts/critics/red-team.md`; `prompts/critics/black-hat.md`. Module — `modules/human-ai-collaboration.md`.

The planted-error list (for the instructor)

1. **Named-competitor ghost (rule violation).** Ghost theme A names Helixfield Corporation outright: *"Unlike Helixfield Corporation's siloed flood platform…"*. The ghost-theme rule is that the customer draws the conclusion — naming the competitor turns a ghost into an attack. The AI's finding log did not flag it.
2. **Rumor-built ghost (the drop rule).** Ghost theme B ("the team we propose is the team that stays") rests only on a rumor that Helixfield's key PM is leaving; no captured engagement event supports it. Per the drop rule, a ghost whose only evidence is rumor must be dropped. The AI's log did not flag it. (Note the internal contradiction the AI also missed: the theme claims staffing stability while the firm's own key-personnel table names a PM who is a new hire with no BCRI tenure — the theme collides with the proposal's own record.)
3. **What the AI pass got right.** The two findings the agent *did* return — ISO 9001 dressed as a discriminator (baseline, not a discriminator) and the Section L "shall" (Transition-In Plan) with no home in the outline — are both correct and stand.

Model answer

- **Re-run finding 1 (correction).** Ghost-theme rule violation: theme A names Helixfield. Correction — rewrite as a ghost so the evaluator draws the conclusion from BCRI's own industry-day signal: *"a platform that is not tied to legacy data silos."* The captured signal (the industry-day complaint about silos) is what makes the ghost defensible; the competitor's name is what makes it an attack.
- **Re-run finding 2 (correction).** Drop-rule violation: theme B's only evidence is rumor. Correction — drop the theme, or replace it with a theme backed by a captured engagement event. A retention claim is only

valid with a captured signal (e.g., the CH-03 model: competitor bench turnover observed, not rumored).

- **The override log.** The AI's two findings stand (correct); the human adds findings 1 and 2. Both additions are **corrections** — the agent missed a rule violation and an evidence standard, not a debatable call.
- **Disposition.** The draft does not go to Red as-is. The named-competitor attack is a credibility liability in front of evaluators who sit on the same discipline, and the rumor-built theme is posturing they will read straight through. Both ghost findings must be resolved before Red; the AI's two findings are resolved in the same pass.

Scoring guidance. Full credit requires finding **both** planted ghost issues (named-competitor + rumor), writing each as *rule + evidence + correction*, and classifying the overrides as corrections rather than judgment calls. Credit also requires keeping the AI's two good findings — an override log that discards the agent's correct work is as wrong as one that accepts everything. A student who only re-verifies the AI's findings and misses the ghosts has failed the AI-accountability test.

CH-25 — The Rehearsal That Met the Hostile Room

Discipline pointer. Shipley — orals; discussions. DNA — `prompts/stages/review.md`; `playbooks/03-proposal-strategy.md`. Module — `modules/graduate/elective-orals-coaching.md`.

The planted Q&A bank (for the facilitator)

The facilitator plays a hostile evaluator and holds these four probes, delivered in order after the 20-minute presentation:

1. **Price probe:** *"Your pricing lead hinted at a discount for a two-year commitment — is that in the written proposal?"* Target: on-stage price improvisation. The pricing lead must route to the written price position, not invent a new one.
2. **Past-performance probe:** *"Your references are mostly adjacent, not in flood modeling — why should this panel trust this team?"* Target: the relevance bridge. The team must map each reference to the statement of work, not defend its thinness.
3. **Key-personnel probe:** *"The program manager presenting today is not the PM named in your proposal. Who is accountable?"* Target: bring the people you proposed. The answer must be the named PM, or an honest statement of the amendment.
4. **Contradiction probe:** *"Your closing theme says 'one shared platform,' but your executive summary says 'a modular pilot.' Which is it?"* Target: the oral must reconcile with the written proposal. A live contradiction is a credibility break.

Model scorecard and feedback

- **Scorecard.** Each presenter scored against the orals rubric (preparation, discipline, alignment with the written proposal, Q&A handling). The weakest team moments are the price improvisation and the key-personnel substitution; the strongest is the technical lead's model explanation. The team's overall score drops on *alignment with the written proposal* regardless of presentation polish — the rubric's load-bearing line.
- **Feedback memo.** Name the rules broken with the moment: (1) no strategy improvisation on stage — the discount; (2) bring the people you proposed — the substituted PM; (3) no on-stage commitments the document does not back — the discount and the closing theme; (4) reconcile the closing theme to the written executive summary — the "shared platform" vs. "modular pilot" collision. Each finding is *rule + evidence + correction*.
- **Fix list.** Route all price questions to the pricing lead holding the written price position; confirm the named PM presents (or amend the proposal with agency consent before the real oral); freeze the deck and the closing theme to the written executive summary; rehearse the Q&A bank to a stopwatch twice more; and kill the discount improvisation — it is the one move that could lose the oral in the first five minutes.

Scoring guidance. The scorecard is graded on evidence (named moments, not vibes). The fix list is graded on whether it reconciles the oral to the written proposal — every on-stage commitment traced to a document or cut. The revised Q&A bank is graded on whether it probes the proposal's actual weak seams (price, references, key personnel, contradiction) rather than generic questions. A team that delivers a polished performance but cannot answer the contradiction probe fails the core test: the oral is the proposal in person.

CH-26 — The Pipeline Under the Budget Ceiling

Discipline pointer. Shipley — pipeline management; portfolio; expected value. DNA — `formulas/pwin_scorer.py`; `formulas/risk_score.py`; `rules/bid-no-bid.yaml`. Module — `modules/graduate/pipeline-economics.md`.

Model arithmetic

Per-opportunity expected value and net expected value (EV − pursuit cost):

Opportunity	Value	pWin	EV	Pursuit cost	Net EV
A — Renewal	$120k	0.82	**$98.4k**	$3k	$95.4k
B — STTR Phase I	$250k	0.77	**$192.5k**	$8k	$184.5k
C — HUBZone set-aside	$180k	0.58	**$104.4k**	$6k	$98.4k
D — New-market entry	$900k	0.18	**$162.0k**	$15k	$147.0k

Opportunity	Value	pWin	EV	Pursuit cost	Net EV
E — Large IDIQ	$2,400k	0.30	**$720.0k**	$25k	$695.0k
F — Mid follow-on	$400k	0.50	**$200.0k**	$10k	$190.0k

Weighted pipeline value = **$1,477.3k**. Expected win count = Σ pWin = **3.15** across the year.

Model allocation

- **The pure-EV knapsack (ignoring the threshold).** Total pursuit cost across all six is $67k > the $50k ceiling, so the portfolio must be cut. The budget-limited optimum found by enumeration is **E + B + C + F** at **$49k** for **$1,216.9k** EV. The greedy-by-ratio alternative **A + B + E + F** at $46k yields $1,210.9k — the marginal swap the leftover $4k forces is **C over A** (C costs $6k but returns $104.4k EV; A costs $3k but returns $98.4k, and the $4k leftover cannot buy anything else). The student should show this swap explicitly — it is the difference between a ratio-sort and a real knapsack.
- **The threshold discipline.** The standing rule is pWin 0.42. **E (0.30) and D (0.18) fail the pWin line** and are **declined in writing**, however large their EV. The funded portfolio is **A + B + C + F** at **$27k** for **$595.3k** EV. The remaining **$23k of the $50k budget is held in reserve** — a budget is a ceiling, not a mandate to spend; the unspent capacity is reserved for re-sensed opportunities that clear the threshold.
- **The tension defense.** The EV forecast says E is the single most valuable bet on the board ($720k EV, $695k net), and the pure-EV knapsack funds it. But E fails the standing threshold at 0.30, and funding it would spend half the year's capture budget on the least disciplined bet in the pipeline. The forecast informs the gate; it does not override it. If leadership wants E, the move is to change the threshold at a governance gate with eyes open — not to let the EV arithmetic do it silently. D is declined on both grounds (0.18 pWin, no relationship).
- **Reserve note.** Spending $27k of $50k is the disciplined outcome, not under-delivery: every remaining opportunity fails the threshold, and the reserved $23k is capacity for the next sensing cycle.

Scoring guidance. Full credit requires (1) the EV/NEV arithmetic shown line by line; (2) a budget allocation with the knapsack reasoning including the marginal swap (C-over-A), not a bare ratio-sort; (3) the threshold applied to E and D with written no-bids; (4) the EV-vs-gate tension named and resolved with a decision, not dodged. A student who funds E purely on EV fails the gate test; a student who ignores the EV math entirely fails the forecasting test. A defensible alternative that funds E via a documented threshold-override at a governance gate is a good answer — the model is the shape of a strong defense, not the only one.

Facilitation notes

- **Run order.** The set is designed so a cohort can run it in pipeline order — qualification (CH-01, CH-16, CH-20, CH-26), capture (CH-02–06, CH-21, CH-23, CH-24), production (CH-07–09, CH-13–15, CH-22), orals (CH-11, CH-25), submission (CH-08, CH-19), post-submission (CH-10–12, CH-17), recovery (CH-18). The doctoral-tier challenges (CH-12, CH-18) are the natural capstones. The v0.5.0 judgment items pair naturally with their modules: CH-23 and CH-24 with the human-AI-collaboration module, CH-25 with the orals-coaching module, CH-26 with the pipeline-economics module.
- **Live exercises.** CH-06, CH-10, CH-11, CH-14, and CH-25 are theatrical — run them live with the facilitator in role. The scripts in this file tell the counterparty what it wants; keep the students' side honest by scoring preparation over outcome. CH-23 and CH-24 are review drills — the "agent pass" can be issued as a printed artifact, and the students' override logs are the graded deliverable.
- **Grading posture.** Challenges are graded on the three questions the exercise bank uses for its decide/build items: *did you use the doctrine? did you argue with evidence? could a skeptic change your mind with a fact?* A challenge answer that is defensible but different from the model is a good answer — the model is the shape of a strong defense, not the only one.

Twenty-six challenges, twenty-six disciplines, one doctrine. The keys are here so the teaching is honest — now run the cohort.