MGT 816 · The Private Firm CFO

S07 — AI-Assisted Forecasting Workshop

Arc 3 · Plan & Measure — the hands-on half of the KPI module

Monday, September 28, 2026 · Prof. John M. Barrios

Before we start:

- ☐ Laptop + desktop Excel open
- ☐ Model Pack downloaded (Canvas → S07 Workshop Kit)
- ☐ Sit with your PS2 group (3–4)
- ☐ One live LLM login at the table
- ☐ No login? Hand up — we re-pair now.

PS2 assigned today — due Mon 10/12, 1:00 pm

Case 2 memo due Wed 9/30, 1:00 pm

Did You Know? Coleco and the Cabbage Patch Kids

DID YOU KNOW?

Ever heard of Coleco Industries? Probably not — but you know the product they brought to market.

- 1983: Coleco launches the Cabbage Patch Kids — the most iconic toy in US history.
- They underestimated demand. Sounds like a good problem, right?
- They had not planned the manufacturing — too little personnel and working capital to supply the surge.
- Empty shelves infuriated parents; the dolls fell out of favor; sales collapsed.
- By 1988 Coleco was paying to meet demand that had vanished. Bankruptcy; rights sold to Hasbro.

© 2026 John M. Barrios · Yale SOM · MGT 816 — for enrolled students only

The demand forecast was almost right. The capacity plan wasn't a forecast at all.

Today's contract: the AI will be wrong — catch it ★

The AI will be confidently wrong at least once in the next hour. Your grade is for catching it.

ANALYST — extracts drivers
blind spot: invents what the case omits

SKEPTIC — benchmarks assumptions
blind spot: cites sources that don't exist

BUILDER — drafts schedule logic
blind spot: misses ramps and lags

BOARD RED-TEAM — attacks the forecast
blind spot: politeness

FLIGHT PLAN

0:12	R1 Analyst
0:24	R2 Skeptic
0:36	R3 Builder
0:52	R4 Scenarios + Red-Team
1:04	Sparring
1:10	One-pager
1:16	PS2

AI never touches the .xlsx — paste in, paste out

Every claim goes in the log: prompt → claim → check → result → change

Every external number sourced or UNVERIFIED

Anatomy of a working prompt ★

SPECIMEN — Prompt 1

"You are an FP&A analyst at an early-stage SaaS company. Below is our planning case, delimited by ###. Extract every quantitative assumption into a table: driver name / value / unit / statement line it feeds / controllable (Y/N); then present the driver tree as an indented list. Do not invent any assumption that is not in the text; end with a list titled OPEN QUESTIONS for anything the case leaves unspecified. ### [paste case_facts.txt] ###"

① ROLE — "You are an FP&A analyst at an early-stage SaaS company."

② CONTEXT — "Below is our planning case, delimited by ###." (paste case_facts.txt)

③ CONSTRAINT — "Do not invent any assumption that is not in the text."

④ OUTPUT FORMAT — a table (driver / value / unit / line / controllable), then the driver tree as an indented list.

⑤ VERIFICATION DEMAND — end with OPEN QUESTIONS for anything the case leaves unspecified.

· MGT 816 · for enrolled students only

Context → role → constraint → format → verification. The protocol is the skill; the vendor is a detail.

Round 1 — Analyst: sketch first, then extract

12 min

ANALYST

1. Paper first (3 min): hand-sketch Redline's driver tree — sales capacity → new logos → active firms → subscription + usage revenue → gross margin → cash. No screens.
2. Run P1 (paste case_facts.txt between the ### delimiters).
3. Reconcile and log it: AI table vs. your sketch vs. the case text.
4. Certify: open drivers.csv — the table of record for the rest of today.

© 2026 John M. Barrios · Yale SOM · MGT 816 — for enrolled students

Deliverable: Verification-log rows for every discrepancy + every OPEN QUESTION.

MANUAL CHECK #1

Extraction fidelity: every AI-listed assumption traced to the case text. Anything the AI "found" that the case never said = an invention. Log it.

Round 2 — Skeptic: benchmark with provenance

12 min

SKEPTIC

1. Run P2 on the certified assumption table (paste it).
2. For every benchmark the AI invokes: trace it to benchmark_sheet.pdf or another checkable source — otherwise stamp the log row UNVERIFIED.
3. Record the AI's "three most damaging assumptions" claim — you test it in Round 4.

Deliverable: Assumption table annotated aggressive / typical / conservative + provenance stamps in the log.

John M. Barrios · Yale SOM · MGT 816 — for enrolled students

MANUAL CHECK #2

Benchmark provenance: every external number sourced or stamped UNVERIFIED. No exceptions — an unsourced "median churn 0.8%/mo" laundered into a driver compounds for 12 months.

0:33 hands poll — whose AI cited a statistic it could not source?

Round 3 — Builder: you own the spreadsheet ★

16 min

BUILDER

1. Run P3 (paste the customer-count assumption rows). The AI returns a 12-column table + formula logic in plain words — never cell references. You implement it.
2. Build the customer-count schedule, then the headcount schedule, in starter.xlsx (labeled rows are waiting).
3. Tie out.

December ending active firms = ⟨shown in class⟩. Not your number? Find the error before continuing — usual suspects: ramp month · churn timing (0% months 1–3, then 1.5%) · the hire-one-month-early rule.

© 2026 · le SOM · MGT 816 · for enrolled students

MANUAL CHECK #3

Roll-forward integrity: beginning + new − churned = ending, all 12 months. Hand-recompute ONE month of revenue from the drivers before trusting the row.

Round 4 — Scenarios + Board Red-Team

12 min

dents only

SCENARIOS (P4)

- Vary ONLY four named drivers: new-logo quota per AE · logo churn rate · documents/firm/month · inference cost/document.
- Base / upside / downside with one-sentence rationales grounded in case facts — no new facts.
- Paste values into your Assumptions sheet. Note the AI's "most sensitive driver" claim.

BOARD RED-TEAM (P5)

- Paste your outputs. Demand the five hardest board questions + an internal-inconsistency hunt:
- revenue vs. sales headcount · CAC vs. marketing spend · support ratio vs. customer count · commissions vs. bookings.

Something in your model is broken. The red-team should find it. Fix it. Log it.

MANUAL CHECK #4

Sensitivity replication: one-variable Excel data table on the AI's claimed most-sensitive driver → confirm or refute in the log. The AI's ranking is a claim, not a fact.

Sparring window: find one more real flaw

6 min

Any prompt you want. One goal: make the AI find one more real flaw in your forecast — beyond the one you just fixed.

"Real" = it changes a number, or an assumption you'd defend differently.

Post to the Canvas thread: your best catch + the exact prompt that produced it.

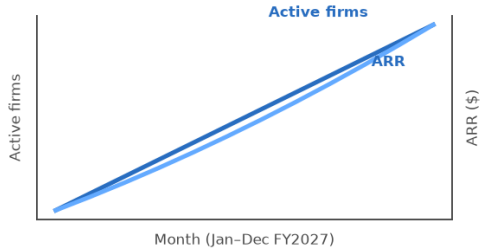
© 2026 John M. Barrios · Yale SOM · MGT 816 — for enrolled students only

This is the only unscripted round. Variance across groups is the point — we compare prompt strategies in 6 minutes.

The board-ready one-pager

6 min + two 90-second screen-shares

① Trajectory



Template placeholder — your model fills this.

② Scenario cash

end-of-FY27 cash + months of runway for base / upside / downside (3×2 table)

③ Top-5 assumptions

value + sensitivity rank from YOUR data table (not the AI's claim)

④ The 6-KPI dashboard row

the six your group chose in S06's pruning exercise, with FY27 exit values (NRR · CAC · LTV:CAC · gross margin · burn multiple · runway, or your defended swaps)

Every number on this page must exist in a model cell. In PS2, a memo number that isn't in the model = –2 points. Each.

Where the AI was wrong today

Hallucinated benchmark

R2's unsourceable stats

Invented assumption

R1's fabrications / OPEN-QUESTION dodges

Arithmetic or logic slip

R3's ramp/lag misses, checksum failures

The planted flaw

support ratio vs. the 1:15 rule — every section's guaranteed row

Captured live in class — see the section board photo posted with this deck.

None of these were hard to catch. All of them were easy to miss. The log is where catching becomes a habit.

PS2: finish what you started ★

PS2 — Redline Labs: the forecast, the memo, and the log · due Mon Oct 12, 1:00 pm · groups of 2–4 (today's groups) · 40 pts — both problem sets count (40+40)

Part	What	Pts
A. Unit-economics warm-up	churn / logo-retention / **NRR** on Redline cohort data (5) + "Spot the AI's errors": half-page AI churn analysis, two planted mistakes — find and fix both (3)	8
B. The build	complete the FY2027 monthly model you started: customer & headcount schedules vs. **5 checksum cells** (6) · income statement + cash walk (6) · **deferred-revenue roll-forward** (3) · model hygiene (3)	18
C. Scenarios + memo	three scenarios **through drivers** (4) · sensitivity table on four named drivers (2) · one-page board memo with the one-pager exhibit (4)	10
AI Verification Appendix	completed log across the Five Checks (2) + two documented AI errors caught, or evidence of checks run (2)	4

AI required · every number human-certified · memo numbers not in the model = –2 each · the final exam (200 pts, closed book) tests this without AI.

Round 5 of the checks happens in PS2: hand-recompute December, top to bottom.

Appendix — reference only, examinable at concept level.

Keep these prompts — they are the durable artifact.

Prompt card I — P1 Analyst · P2 Skeptic

P1 — Driver identification (Analyst):

"You are an FP&A analyst at an early-stage SaaS company. Below is our planning case, delimited by ###. Extract every quantitative assumption into a table with columns: driver name | value | unit | financial statement line it feeds | controllable by management (Y/N). Then present the driver tree from sales capacity to ending ARR as an indented list. Do not invent any assumption that is not in the text; instead, end with a list titled 'OPEN QUESTIONS' for anything the case leaves unspecified. ### [paste case_facts.txt] ###"

P2 — Assumption benchmarking (Skeptic):

"You are a skeptical FP&A reviewer. Here is our assumption table: [paste]. For each assumption, classify it as aggressive / typical / conservative for a Series-A B2B SaaS company in 2026, with one sentence of reasoning. Wherever you invoke an external benchmark, name the source and your confidence in it, and tag the line 'VERIFY BEFORE USE.' Finish by naming the three assumptions that, if wrong, do the most damage to year-end cash, and why."

© 2026 John M. Barrios · Yale SOM · MGT 816 — for enrolled students only

Prompt card II — P3 Builder · P4 Scenarios

P3 — Schedule drafting (Builder):

"Build a monthly FY2027 customer-count schedule from these assumptions: [paste relevant rows]. Output (a) a 12-column table with beginning balance, new logos, churned logos, ending balance; and (b) for each row, the formula logic in plain words so I can implement it in Excel myself. Explicitly flag any month where a ramp, training lag, or hiring rule changes the calculation. Do not give me Excel formulas with cell references — I own the spreadsheet."

P4 — Scenario building:

"Create base, upside, and downside scenarios by varying ONLY these four drivers: [list]. For each scenario, give the driver values and a one-sentence rationale grounded in the case facts — no new facts. Present as a table I can paste into my assumptions sheet. Then state which single driver you believe moves December cash the most; I will test your claim with a data table and report back."

© 2026 John M. Barrios · Yale SOM · MGT 816 — for enrolled students only

Prompt card III — P5 Board Red-Team · P6 Board Narrative

P5 — Forecast critique (Board Red-Team):

"You are a board director with a venture background reviewing a Series-A company's annual operating plan. Here are our forecast outputs: [paste one-pager numbers]. Ask the five hardest questions you would ask in the board meeting. Then identify any internal inconsistencies — revenue growth vs. sales headcount, CAC vs. marketing spend, support ratio vs. customer count, commission expense vs. bookings. Be direct; do not soften."

P6 — Board narrative (Drafter — used in PS2 Part C, not today):

"Draft a 250-word board narrative from ONLY the following verified numbers: [paste]. Structure: the headline, the base case, the single biggest risk, and the pre-agreed trigger that would cause us to re-forecast. If a sentence would require a number I did not give you, leave a bracketed placeholder instead of inventing one."

© 2026 John M. Barrios · Yale SOM · MGT 816 — for enrolled students only

P6 is used in PS2 Part C — not today.

The Five Manual Checks ★

1. **Extraction fidelity** — every AI-listed assumption traced to the case text.
2. **Roll-forward integrity** — beginning + additions − churn = ending, all 12 months; cash walk ties to net income ± working capital.
3. **Benchmark provenance** — every external number sourced or stamped UNVERIFIED.
4. **Sensitivity replication** — a student-run data table confirms or refutes the AI's "most sensitive driver" claim.
5. **Column recompute** — hand-recompute December, top to bottom (PS2).

© 2026 John M. Barrios · Yale SOM · MGT 816 — for enrolled students only

Each check must appear in your Verification Log. The log is graded (4 pts); the judgment applied to AI output cannot be outsourced.

Fallback: critique the canned transcript

AI access down? The lesson isn't.

- Open Canvas → S07 Workshop Kit → Fallback → the transcript for your current round. It is a real AI response to the round's prompt, pre-generated on this exact pack.
- Run the round's manual check against the transcript instead of a live AI: mark every claim Confirmed / Corrected / Rejected in your log, exactly as you would live.

© 2026 John M. Barrios · Yale SOM · MGT 816 · for enrolled students only

The transcripts contain the same species of errors a live model makes — at least one per round worth catching. The verification skill survives a total outage.