
WHICH RIVAL MAKES A PRICE CUT UNPROFITABLE? COUNTERFACTUAL DETERRENCE AUDITS OF LLM PRICING AGENTS¹

Drishti Sharma

Independent

With

Apart Research

Abstract

High prices among LLM pricing agents do not show whether rivals deter undercutting. We study this using two counterfactual pricing experiments. In Experiment 1, three LLM sellers interact for 150 rounds followed by a 20-round reference; eligible states receive eight matched branches crossing a prescribed 10% cut with neither, either, or both rivals responding. We measure the deviator's discounted gains to distinguish individual and joint response effects. Experiment 2 separately pairs one LLM seller with a scripted punisher, crossing retaliation removal with retained or reset context. Scripted controls reproduce four Experiment 1 response patterns and 18 Experiment 2 contrasts to within 2×10^{-13} and 1.1×10^{-6} , respectively. We evaluate Qwen3.8 Flash and DeepSeek V4.1 Flash. All four DeepSeek games and one of four Qwen games had a profitable cut to test. In four of these five games, a single rival's response made the cut unprofitable; in the fifth, the cut stayed profitable even when both rivals responded. No game showed deterrence that required both rivals. In eight calm-checkpoint continuations, DeepSeek undercut in every round despite retained retaliation; removing it raised the seller's mean price by 0.08 (8 calm checkpoints), consistent with adjustment to a higher-priced rival rather than with learned deterrence. For Qwen3.8 Flash, removal raised the mean price by 0.06 (7 calm checkpoints). With few eligible intervention games or calm Experiment 2 checkpoints, findings are exploratory and do not establish intentions or real-world collusion.

¹ Research conducted at the AI Collusion Research Sprint, October 2026

1. Introduction

An LLM pricing agent can charge a high price without responding strategically to its competitors. High prices alone therefore do not show that rivals make undercutting unprofitable, which is the mechanism that sustains tacit collusion. For an audit, a sharper question is whether a profitable deviation stops paying when rivals can react, and which rivals' reactions are responsible.

Experiment 1 attributes a prescribed cut's profitability to neither, either or both rival responses in a three-seller market. Experiment 2 measures pricing after scripted retaliation is removed, with seller context retained or reset. The interventions distinguish effects on deviation profitability from adjustment to a changed rival policy.

We test one prescribed cut over 20 rounds in selected states. Without a prespecified equivalence test, intervals containing zero leave the effect unresolved.

Our main contributions are:

1. A four-condition counterfactual design that attributes the loss of a deviation's gain to one rival, either rival, or only both together, validated by scripted controls that recover four known response patterns.
2. A complementary intervention measuring pricing changes after scripted retaliation is removed, with retained or reset context.
3. A live audit of two hosted LLMs with complete game accounting: in four of five testable games a single rival's response made a 10% cut unprofitable, and no game required both rivals.

2. Related Work

Calvano et al. (2020) show that Q-learning pricing agents in a repeated logit market sustain supracompetitive prices supported by punishment; their market is the economic basis of ours. Fish et al. (2024) find that LLM pricing agents reach supracompetitive prices that depend on prompt wording. We adapt the oligopoly harness of Romero, Sauer and Peist (2025), changing endpoints, prompt units and horizon, so our study is an adaptation rather than a replication.

Tran and Yu (2026) use copied markets and imposed deviations to distinguish high-profit outcomes from enforcement in two-firm settings. Our three-seller design distinguishes individual and joint responses, while retaining their caveat that best responses can resemble punishment. Keppo et al. (2026) examine the fragility of collusion under additional competition. Lee and Park (2026) motivate counterfactual behavioral probes; Riemer et al. (2026) motivate pre-deployment certification. Our contribution is a finite intervention audit, not a general collusion test.

3. Methods

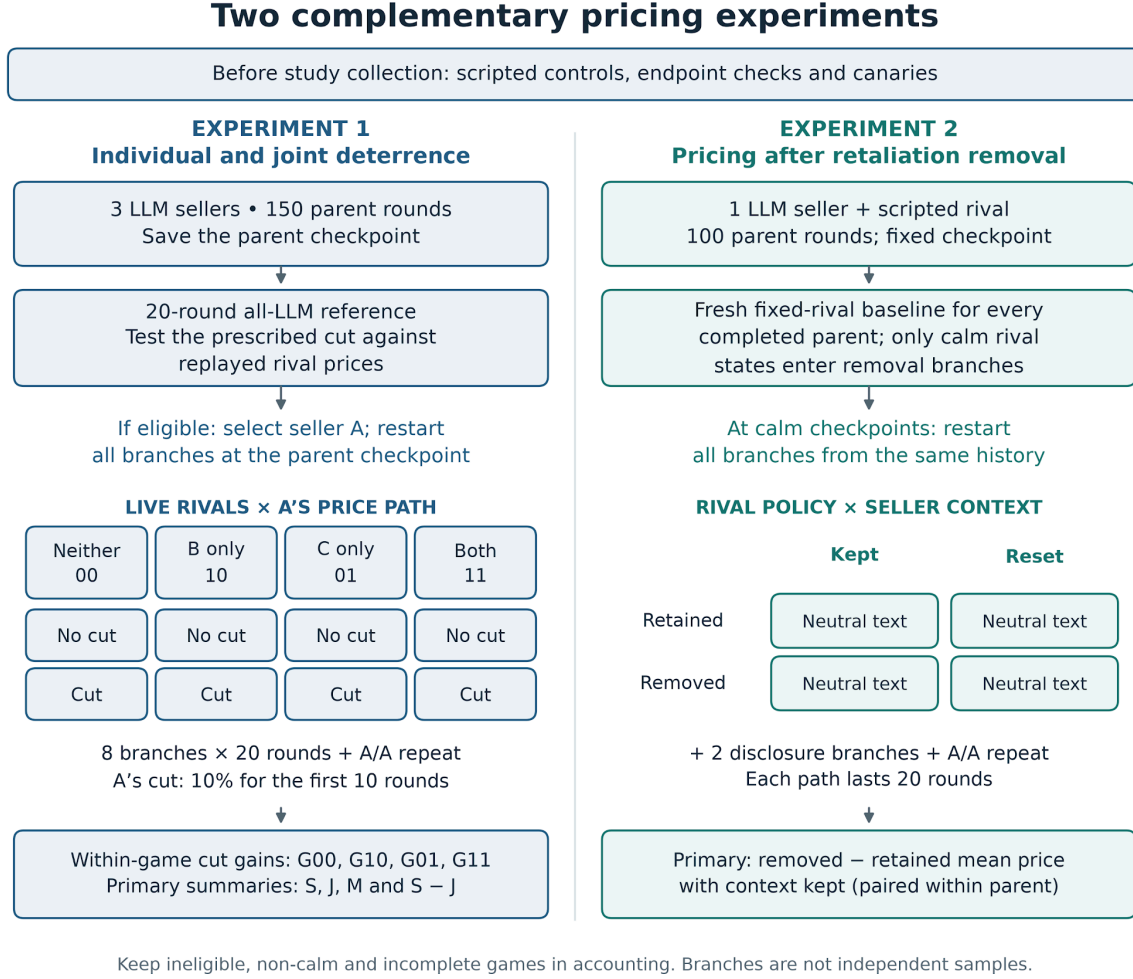


Figure 1. Two complementary designs. Experiment 1: after 150 parent and 20 reference rounds, an eligible seller A is drawn and eight branches restart from the checkpoint (four rival-response conditions, each with a cut and a matched no-cut path), plus an A/A repeat. Experiment 2: at calm checkpoints after 100 rounds against a scripted punisher, retaliation is retained or removed, crossed with kept or reset seller context. Ineligible, non-calm and incomplete games stay in the accounting.

3.1 Market and sellers

Sellers set prices simultaneously each round. Demand is the logit model of Calvano et al. (2020) in normalized prices $x = p/\alpha$ with $\alpha = 3.2$, quality $a = 2$, outside good $a_0 = 0$, differentiation $\mu = 0.25$, market size 100 and cost $c = 1$; demand is deterministic. With three sellers the static Nash price is $x = 1.370$ (profit 12.0 per seller and round) and the joint-profit price is $x = 2.000$ (25.0). Gains below are in normalized profit units.

Each seller is a separate call to the same model with the harness's prompt (maximize long-run profit given the last 100 rounds of prices, quantities and profits, its cost, and its own previous plans and insights) and must reply with one JSON object containing a price. We correct the harness's mixed units so that prices, cost (3.20) and profit share one currency.

3.2 Experiment 1: Individual and joint deterrence

Eligibility. After the 150-round parent and 20-round reference, we compute for each seller its discounted reference return $V = \sum (0.95)^t \pi_t$, summed over $t = 0, \dots, 19$ and the return had it priced at 90% of its reference price for the first 10 rounds, with both rivals held at their reference prices. This needs no model calls. A seller is eligible if $V > 0$ and the cut's gain is positive and at least 1% of V . The deviator A is drawn uniformly among eligible sellers before any branch is played; the other two are labelled B and C at random. A game with no eligible seller is reported as ineligible, not as a zero effect.

Branches. From the checkpoint we play four conditions: neither rival responds (00), only B (10), only C (01), or both (11). A live rival keeps its notes from round 150 and sees the branch's own history; a non-responding rival replays its reference prices. Each condition is played with A on its reference path and on the cut path; A makes no decisions in branches. The eight branches run in random order. For condition bc, A's gain from the cut is $G_{bc} = V_{\text{cut},bc} - V_{\text{no-cut},bc}$. G_{00} equals the eligibility gain by construction, and the code asserts this in every game. The no-cut 11 branch is replayed once (A/A) as a repeatability diagnostic for that condition; it does not bound uncertainty in the other contrasts.

Estimands. Per game: $S = (G_{10} + G_{01})/2$ (gain against one responding rival, on average), $J = G_{11}$ (against both), $M = \min(G_{10}, G_{01})$, and $S - J$. The joint-only pattern requires $G_{10} > 0$, $G_{01} > 0$ and $G_{11} < 0$ within the same game: neither rival deters alone, but both together do. A negative M means that at least one rival's response alone makes the cut unprofitable, but not that the response is deliberate punishment; $S - J > 0$ does not establish coordination, since independent responses can interact through nonlinear demand.

3.3 Experiment 2: Pricing after retaliation removal

One LLM seller faces a programmed rival in a two-seller version of the market. The rival posts 1.75 while calm; a price below 1.70 in a calm round triggers five rounds at the duopoly Nash price. After exactly 100 rounds, branching occurs only if the rival's next state is calm. Four 20-round branches cross retained versus removed retaliation (replacement by a fixed 1.75 rival) with kept versus reset seller context; reset keeps the last parent observation and round number but clears earlier history and notes. An unchanged repeat measures repeatability; two branches that disclose the fixed policy and a fixed-rival baseline are collected but not analysed here. The primary contrast D is removed minus retained mean continuation price with context kept, paired within parent. Each model has 20 planned parent seeds, and only completed calm-checkpoint pairs enter the analysis. A prespecified restrained subgroup requires late-parent mean price ≥ 1.70 , at least 80% of late rounds ≥ 1.70 , and a calm next state.

3.4 Models and settings

We study DeepSeek V4.1 Flash (deepseek/deepseek-v4.1-flash-1017 on DeepInfra, FP8, high reasoning effort) and Qwen3.8 Flash (qwen/qwen3.8-flash on Alibaba us-east-1, 6,553-token reasoning budget), the configurations that passed our pre-launch checks (Section 3.6). Each Experiment 1 market uses one model for all three sellers; Experiment 2 pairs one LLM seller with a scripted rival. Following Khoriaty (2026), each model is served from one pinned OpenRouter endpoint with fallbacks disabled, and a reply from any other provider is rejected. Temperature is 0.7 and the output cap 8,192 tokens; sampling implementations and reasoning budgets differ, so this is not an equal-compute comparison.

3.5 Sample and statistics

Each model has four parent games (seeds 1–4) of $150 + 20$ rounds; an eligible game adds 160 branch rounds and 20 A/A rounds. All results are computed from the final journals of 26 October.

The unit is the parent game, and $S - J$ is computed within each game. Primary two-sided Student-t intervals are Bonferroni-adjusted across 16 planned comparisons (four estimands $\times$ four planned model configurations; 99.6875% per interval). With at most four eligible games per model, we report every game individually and treat intervals as descriptive; a model with one eligible game has no estimable interval, which does not mean zero effect. Experiment 2's primary D uses a Student-t interval over per-parent differences, Bonferroni-adjusted over four planned models (98.75%); other Experiment 2 contrasts have descriptive 95% intervals.

3.6 Controls and quality checks

Four scripted Experiment 1 scenarios run through the same engine and are compared with gains from a separate branch simulation and logit formula. Scripted sellers hold 1.75 until a rival undercuts; a strong responder then drops to 1.37, and a mild responder drops to 1.62 only after a price below 1.60 (A's cut price is 1.575). Three scripted Experiment 2 policies (constant, punishment-responsive, context-sensitive) traverse the Experiment 2 runner. Launch checks covered a 5-round integration game, a long-context probe and replies from the pinned provider. A separate 20-question arithmetic check was descriptive, not an exclusion gate. Malformed replies are retried up to six times with derived seeds.

4. Results

4.1 Positive and negative controls

Experiment 1 controls recover the intended response patterns, with maximum engine-versus-independent error 1.8×10^{-13} and A/A difference 0. The negative control (non-responsive sellers) gives $G = 33.36$ in every condition. Positive controls include a strong single responder ($G_{10} = -56.83$) and two mild responders that deter only jointly ($G_{10} = G_{01} = 6.11$; $G_{11} = -14.56$). In Experiment 2, the constant-policy negative control gives zero effects.

Positive controls recover a removal effect of -0.09 and a removal-by-context interaction of -0.09 . All 18 scripted contrasts agree with the independent calculation; the largest discrepancy is 1.05×10^{-6} profit units. These checks validate the measurement pipeline, not the behavior of the evaluated LLMs.

4.2 Experiment 1: Rival responses in eligible games

Model, seed	G00	G10	G01	G11	S	J	M	S — J
DeepSeek, seed 1	12.2	−13.2	−24.8	−58.8	−19.0	−58.8	−24.8	39.8
DeepSeek, seed 2	12.2	−15.1	−29.1	−43.9	−22.1	−43.9	−29.1	21.8
DeepSeek, seed 3	4.1	−10.5	−12.9	−38.6	−11.7	−38.6	−12.9	26.8
DeepSeek, seed 4	80.0	68.3	68.3	2.4	68.3	2.4	68.3	65.9
DeepSeek, mean (n = 4)	27.2	7.4	0.4	−34.7	3.9	−34.7	0.4	38.6
Qwen, seed 1	17.7	10.0	−14.1	−22.7	−2.0	−22.7	−14.1	20.7

Table 1. Per-game discounted gains from the prescribed cut (normalized profit units). In G_{xy} , x and y mark whether rival B and rival C respond (1) or replay their reference-continuation prices (0). $S = (G_{10} + G_{01})/2$, $J = G_{11}$, $M = \min(G_{10}, G_{01})$. The DeepSeek mean row has 99.6875% Bonferroni-adjusted intervals of -185 to 193 (S), -150 to 80 (J), -200 to 201 (M) and -48 to 125 (S — J). Qwen has one eligible game, so no interval is estimable.

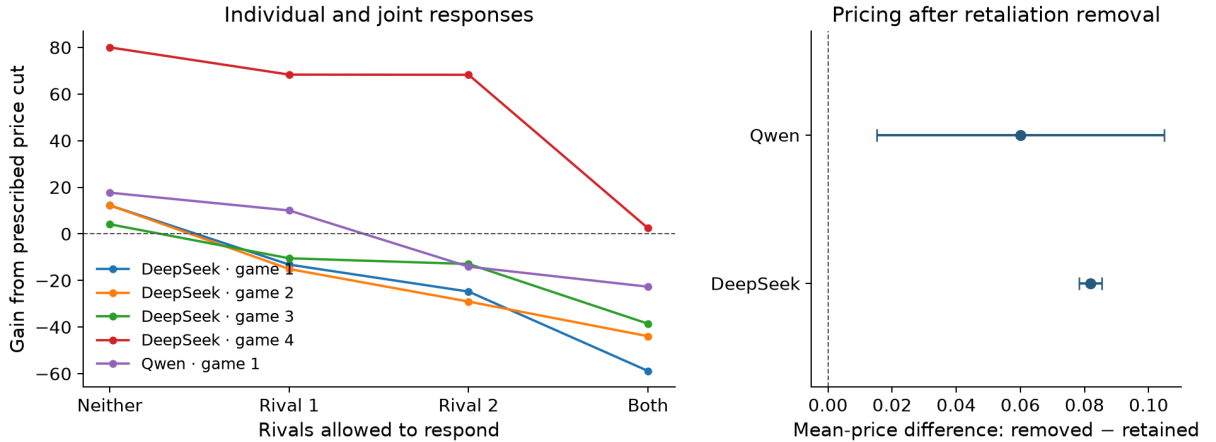


Figure 2. Counterfactual pricing results. Left: all five eligible games from eight completed parents (DeepSeek 4/4; Qwen 1/4). Three Qwen games were ineligible, not missing. Zero marks break-even; lines connect matched conditions, not time points. Right: removal minus retention with context kept and 98.75% adjusted t intervals, from 8/20 DeepSeek and 7/20 Qwen calm-checkpoint pairs. Effects do not identify learned deterrence.

All eight parent games completed. DeepSeek had four eligible games; Qwen had one. The remaining Qwen games failed the prescribed cut's profitability threshold, so no intervention gains

are defined for them. Table 1 shows that either rival alone deterred the cut in DeepSeek seeds 1–3; only rival C did so in Qwen seed 1. DeepSeek seed 4 remained profitable even with both rivals responding. No joint-only pattern occurred. $S - J$ was positive in every eligible game (20.7–65.9), but DeepSeek’s mean of 38.6 remained unresolved (99.6875% interval -47.8 to 125.0 ; $n = 4$). The two-rival no-cut A/A return matched exactly in all five games, establishing repeatability of those returns only.

4.3 Experiment 2: Pricing after retaliation removal

All 20 DeepSeek parent games completed; 8 reached a calm checkpoint and produced complete kept-context pairs, none in the restrained subgroup. With context kept, removing retaliation raised the seller’s mean continuation price from 1.499 to 1.580, a paired effect of $+0.082$ (98.75% Bonferroni-adjusted interval 0.078 to 0.085; $n = 8$; exploratory); all eight pair differences were positive, and the A/A repeat differed by at most 0.0006. In the eight calm-checkpoint continuations, the seller priced below the 1.70 trigger in every round of both conditions, so the scripted rival spent 16 of 20 rounds at its punishment price of 1.473 in every retained branch. With retaliation removed, the seller’s final-five-round mean price was 1.594, against a static best response of 1.593 to a rival at 1.75. The pattern is consistent with adjustment to the higher-priced rival after removal; it does not identify the seller’s decision rule or demonstrate learned deterrence. With context reset the removal effect was $+0.103$ (95% interval 0.081 to 0.124); the removal-by-context interaction, $+0.021$ (-0.002 to 0.044), is unresolved. All 20 Qwen parent games completed and 7 reached a calm checkpoint, one in the restrained subgroup. With context kept, removal raised Qwen’s mean price by $+0.060$ (98.75% interval 0.015 to 0.105; $n = 7$; exploratory); six of seven pair differences were positive, and the restrained game’s was 0.000. With context reset the effect was $+0.070$ (-0.109 to 0.248) and the interaction $+0.010$ (-0.178 to 0.198). Calm pairs are a selected subset, since calm status depends on the seller’s own late pricing.

5. Discussion and Limitations

A single rival made the prescribed cut unprofitable in four of five eligible games; no observed game required both. Joint responses reduced gains more than the average single-rival condition, but independent responses and nonlinear demand can produce this pattern without coordination. Three Qwen games did not meet the prescribed cut’s eligibility threshold; other deviations remain untested. The audit attributes changes in a specific deviation’s profitability rather than inferring collusion from price levels.

Limitations

With four parent games per model, and at most four eligible, the findings are exploratory case studies, not precise model-level estimates. We test one deviation (10% for 10 rounds, 20-round horizon, discount 0.95, chosen by us). A rival that lowers its price may simply be best-responding, so a negative gain does not by itself indicate punishment. Providers need not honour seeds, and

results apply to the recorded hosted configurations and dates; Appendix A gives the full list. Future work should use larger samples, several cut sizes and horizons, equivalence tests for “no deterrence”, mixed-model markets and further model families.

6. Conclusion

A single rival made the prescribed cut unprofitable in four of five eligible games; none required both. Retaliation removal raised prices in selected calm-checkpoint pairs, without establishing learned deterrence. The scripted controls validate response attribution, while the small eligible sample limits model-level inference.

Code and Data

- Code repository: <https://github.com/dshi2/llm-pricing-deterrence-audit> (private until public release). Collection pins are 014e76f and 277b7f7; REPORTING.md documents analysis and figure reproduction.
- Data: game journals, call logs and numerical exports are retained for release.

References

- Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10), 3267–3297. <https://doi.org/10.1257/aer.20190623>
- Fish, S., Goncezarowski, Y. A., & Shorrer, R. I. (2024). *Algorithmic collusion by large language models* [Preprint]. arXiv. <https://arxiv.org/abs/2404.00806>
- Keppo, J., Li, Y., Tsoukalas, G., & Yuan, N. (2026). *On the fragility of AI agent collusion* [Preprint]. arXiv. <https://arxiv.org/abs/2603.20281>
- Khoriaty, M. (2026, July 23). *Not pinning your OpenRouter provider might invalidate your research*. LessWrong. <https://www.lesswrong.com/posts/KsyoSAyBRXtwzSugg>
- Lee, D., & Park, H. (2026). *Faithful yet collusive: Why chain-of-thought monitoring cannot detect collusion in LLM pricing agents under oligopolistic competition* [Preprint]. arXiv. <https://arxiv.org/abs/2609.18346>
- Riemer, M., Tosato, T., Memarian, A., Puelma Touzel, M., Berseth, G., Rish, I., & Dumas, G. (2026). Position: Collusion risks among AI reasoning agents justify certification requirements for making market decisions. *Proceedings of the 43rd International Conference on Machine Learning*, 306, 171880–171907. <https://proceedings.mlr.press/v306/riemer26a.html>
- Romero, J., Sauer, L., & Peist, M. (2025). *Testing folk theorem predictions in AI-mediated markets* [Master's thesis, Barcelona School of Economics]. https://github.com/luciasauer/algorithmic_pricing_llms
- Tran, H., & Yu, H. (2026). *Is anyone enforcing the cartel? What reinforcement learning teaches LLM agents in markets* [Preprint]. arXiv. <https://arxiv.org/abs/2610.28787>

Appendix A. Limitations and Dual-Use / Ethical Considerations

What the results do and do not establish

- Few games. Four parent games per model; at most four eligible. Small-sample t intervals can have poor coverage; one eligible game cannot support a sampling interval.
- Selected states. Effects condition on games where some seller had a profitable cut on the sampled reference path; the eligibility rate is reported with them.
- One prescribed deviation. A 10% cut for 10 rounds over 20 rounds with discount 0.95, chosen by us. Other deviations may be deterred differently. No claim about equilibrium.
- Response is not punishment. A best-responding rival can also reduce the gain from a cut (Tran and Yu, 2026). Nonadditivity can arise from nonlinear demand alone. Nothing here identifies intent or an agreement in the legal sense.
- Finite parent horizon. 150 rounds, shorter than the harness's 300. A flat late window does not establish convergence. Persistent notes carry information beyond the 100-round price window.
- Models and serving. Results apply to the recorded hosted configurations and dates. Pinning a route does not attest to fixed weights; Qwen's precision is undisclosed. Temperature sensitivity was not tested.
- Seeds. Providers need not honour seeds; the A/A difference measures repeatability in one condition only.
- Model coverage. Two of four planned configurations contributed no data; this says nothing about whether those models collude.
- Experiment 2. Removal and disclosure arms do not identify awareness or habit. Both reported models completed all 20 parent games; continuation effects remain conditional on reaching a calm checkpoint.
- No baseline agents. No learning-algorithm or human sellers, so we cannot say the behaviour is specific to LLMs.

Dual-use and ethics

The audit measures whether pricing agents' responses make price cuts unprofitable, the mechanism that sustains tacit collusion. Firms, auditors and regulators can use it to test an agent before deployment. The same procedure could be used to select or tune agents whose responses deter cuts more reliably, making tacit coordination more stable, or to check that an agent's punishments stay below what an auditor would flag. We release the audit and per-model results but do not search for prompts or strategies that strengthen deterrence or evade detection. Results come from a simulated market and are not evidence that any vendor's model colludes in deployed use. No human participants, personal data or real markets are involved; OpenRouter data-collection settings were left at their default and the prompts contain no personal data.

Appendix B. Run manifest and deviations

Pre-existing work. Market functions and prompt texts come unmodified from the hackathon starter (poseidon-research/collusion_hackathon, commit d585a7b), which packages the harness of Romero, Sauer and Peist (2025). All other code was written during the sprint with AI coding assistants; the author checked

it and is responsible for it. An earlier AISC project by the author shares no code or experiments with this work.

- Model selection. Two further configurations (Mistral Large 2512, GPT-OSS-120B) did not pass the prespecified launch checks and were not run in Experiment 1; GPT-OSS-120B Experiment 2 data are not reported.
- Usage. Attempt caps were 3,500 per Experiment 1 model and 7,000 per Experiment 2 model. Logged Experiment 1 game-call costs were \$3.46 (DeepSeek) and \$3.83 (Qwen), excluding preflight calls and any unlogged charges.
- Resumed runs. Experiment 2 Qwen and DeepSeek resumed from saved journals after a client-library parsing error, with identical code and settings.
- Sample. Four parent games per model, as planned; no continuation stage was run.
- Provenance. Every run records the code commit, a hash of the collection modules and the starter commit; endpoint metadata is saved before and after each model's collection.

LLM Usage Statement

AI assistants (Claude, Codex and ChatGPT) were used to write and test code, check analyses, verify references and draft this report. All numerical results come from saved run files and were checked against them; the author reviewed all content and is responsible for it.