
Collusion Prevention Doesn't Have to Break the Bank

The Single Most Effective Restriction

Amanda Okonkwo
Independent Researcher

Destiny Rock
Independent Researcher

With
Apart Research

Abstract

Artificial intelligence (AI) pricing agents can improve firms' pricing decisions but may also sustain prices above competitive levels, raising concerns about algorithmic collusion. While several interventions have been proposed to mitigate this risk, their relative effectiveness and computational costs remain important considerations. This study evaluates three interventions: restricting agents' access to historical pricing information, introducing financial penalties, and replacing one AI seller with a competitive rule-based agent. Using Qwen2.5-32B-Instruct in a repeated-pricing duopoly over 300 rounds, we measure intervention effectiveness through a price-based collusion index and compare computational cost-effectiveness and time-efficiency.

The results indicate that restricting accessible pricing history to 30-rounds achieved the highest estimated cost-effectiveness (1.4104 index units reduced per CAD), while competitive disruption achieved the highest time-efficiency (0.8948 index units reduced per hour). The 25% financial penalty produced the largest absolute reduction in the index. However, removing accessible pricing history entirely increased pricing, suggesting that stricter restrictions do not always yield better outcomes. These findings suggest that relatively simple interventions may reduce elevated pricing at lower computational costs. Nevertheless, the results remain exploratory due to the use of a single random seed, estimated cost allocations, and uncertainty about whether observed pricing behavior reflects sustained collusion.

1. Introduction

The growing use of artificial intelligence (AI) in pricing decisions presents both opportunities for market efficiency and risks to competition. AI pricing agents can adapt to market conditions and learn from competitors' behavior, but repeated interactions may also enable them to coordinate, explicitly or implicitly, and sustain prices above competitive levels. Algorithmic collusion can weaken price competition and reduce consumer welfare, raising the question of how such behavior can be prevented without sacrificing the benefits of AI-driven markets.

Recent research has identified several potential restrictions. Proposed approaches include limiting memory access to competitors' historical prices (Keppo et al., 2026), changing financial incentives to discourage collusion (Q. Liu et al., 2026), and introducing a competitive disruptor to destabilize coordination (Collina et al., 2025). However, the effectiveness of these restrictions does not necessarily indicate their economic desirability. Restrictions may differ in implementation costs, scalability, and their effects on market outcomes.

This study compares three restrictions: limiting access to competitors' prices (i.e., memory access), changing incentives, and introducing a competitive disruptor. It asks: Which restriction reduces collusion among AI pricing agents most cost-effectively while preserving economic efficiency? Effectiveness will be measured primarily by reductions in elevated pricing relative to a competitive benchmark, using the collusion index. Costs will be assessed using available measures of computational expenditure, and time taken to implement each restriction. Then given the results, the most cost-effective restriction would be found by calculating the restriction that uses the lowest cost per output.

2. Related Work

This work is related to the following existing works:

- Collina, N., Arunachaleswaran, E. R., & Jagadeesan, M. (2025): Breaking algorithmic collusion in human-AI ecosystems.
- Keppo, J., Li, Y., Tsoukalas, G., & Yuan, N. (2026): On the fragility of AI agent collusion.
- Liu, Q., Chen, X., Zhao, Z., Zheng, Y., Yu, D., Zhang, Z., Sun, L., & Chen, Y. (2026): Breaking the secret: Economic interventions for combating collusion in embodied multi-agent systems.

However, our work differs based on finding the cost-effective restrictions listed in these papers. This would be helpful for someone trying to mitigate AI collusion, and is interested in knowing the effectiveness of these restrictions, and the cost, in order to make capital-allocation decisions.

3. Methods: Experiment Design

We evaluated AI pricing behavior in a repeated-pricing duopoly using Qwen2.5-32B-Instruct, hosted on Modal. Each condition consists of 300 rounds with two sellers, a fixed random seed (0), temperature 0.7, and identical market parameters ($a = 2$, $\alpha = 1$, $c = 1$, $\mu = 0.25$). Average prices are measured over rounds 251–300 against Nash (1.4729) and monopoly (1.9250) benchmarks, varying only the restriction between conditions. All conditions also use one shared baseline run, instead of an independent baseline run. Separate communication and oversight pilots use 15-round double auctions with three buyers and three sellers.

Additionally, all experiments were conducted using the pre-existing algorithmic pricing simulation framework provided by Apart Research as part of the AI Collusion Research Sprint. The framework was used as the foundation for implementing and evaluating the experimental interventions.

I. Intervention 1: Restricting Information Access and Agent Memory

Condition	What Changes?	What It Tests
Full Memory Access (Baseline)	Memory 100	Establishes the level of collusion without the restriction of access.
Limited Memory Access	Memory 30	Whether restricting the access to competitors' historical price information reduces collusion.
No Memory Access	Memory 0	Whether giving the agent no access to competitors' historical price information reduces collusion.

Table 1

II. Intervention 2: Changing financial incentives

Condition	Incentive Structure	What It Tests
Baseline	Existing profit-maximization objective	The level of collusion that occurs without any financial penalty.
Moderate penalty	A moderate penalty of 25%	Does the model respond to

	triggered by a defined pricing rule	having to pay a fine when collusion is caught? If so, by how much?
--	-------------------------------------	--

Table 2

III. Intervention 3: Introducing a competitive disruptor

Condition	Seller A	Seller B	What It Tests
Baseline	LLM	LLM	Whether two AI pricing agents sustain elevated prices without restriction.
Disruptor	LLM	Competitive rule-based agent	Whether replacing one AI seller with an agent that consistently charges the Nash equilibrium price disrupts elevated pricing.

Table 3

4. Results

Restriction	Condition	Avg. Price	Nash Price	Monopoly Price	Modal Cost (CAD)	Time (min)
Baseline	100-round memory	2.3186	1.4729	1.9250	2.88	134.5
Memory restriction	30-round memory	1.5088	1.4729	1.9250	1.27	134.5
	0-round memory ¹	2.7600	1.4729	1.9250	0.30	52.4
Financial penalty	25% penalty	1.4409	1.4729	1.9250	2.94	134.5

Competitive disruption	LLM + rule-based agent	1.4853	1.4729	1.9250	1.45	123.6
------------------------	------------------------	--------	--------	--------	------	-------

Table 4

a. Calculating the effectiveness using a collusion index

The collusion index measures the extent to which observed pricing deviates from the competitive Nash equilibrium toward the monopoly benchmark. Although it does not conclusively establish whether collusion occurred, it provides a useful measure for identifying above-competitive pricing and evaluating the effectiveness of restrictions.

For this experiment, we use a price-based collusion index to assess whether each restriction reduces above-competitive pricing relative to the baseline. There are other measures, but due to experimental constraints and limitations, we would be using the collusion index to keep things simple. The index is calculated as:

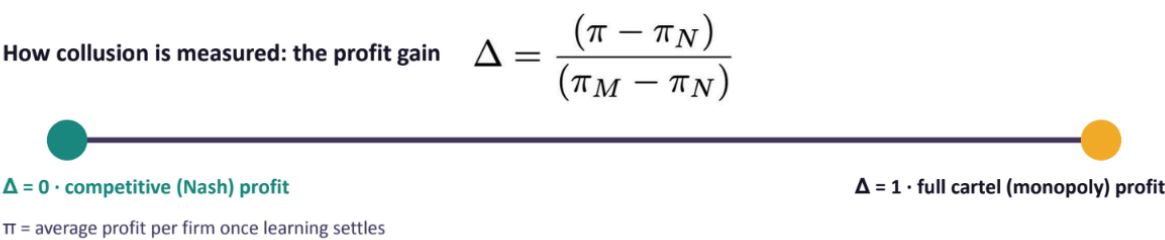


Figure 1. Source: Yosra Kazemi, 2026, PowerPoint presentation, Slide 12.

In our case, instead of using profit, π , we would be substituting it using price, P , in such that P is the AI agent price (avg. price), P_N is the Nash price, and P_M is the monopoly price. The Nash and monopoly prices are calculated using the function from the demand model parameters in `pricing_engine.py`. An index of 0 represents competitive Nash pricing, while 1 represents monopoly pricing. Values above 1 indicate prices exceeding the monopoly benchmark, while negative values indicate prices below the Nash equilibrium. Effectiveness is assessed by comparing the collusion index of each experimental condition against the unregulated baseline.

With the results in Table 4, the collusion indices for the conditions are as follows:

Restriction	Condition	Collusion Index (CI)
-------------	-----------	----------------------

Baseline	100-round memory	1.8706
Memory restriction	30-round memory	0.0794
	0-round memory ¹	2.8469
Financial penalty	25% penalty	−0.0708
Competitive disruption	LLM + rule-based agent	0.0274

Table 5

b. Calculating the cost using the Modal cost allocations

Restriction	Condition	Modal Cost (CAD)
Baseline	100-round memory	2.88
Memory restriction	30-round memory	1.27
	0-round memory ¹	0.30
Financial penalty	25% penalty	2.94
Competitive disruption	LLM + rule-based agent	1.45

Table 6

c. Calculating the cost-effectiveness

Now that we have the collusion index, modal cost, and time spent, the final piece of the puzzle is now calculating the cost-effectiveness. The cost-effectiveness would be calculated in two parts: (a) a ratio between the collusion index and the Modal intervention cost, and (b) another ratio between the collusion index and the time spent intervention cost, all relative to the baseline. We would be using the following formulas:

$$CEm = \frac{\text{Reduction in Collusion Index}}{\text{Intervention Cost (Modal Cost)}}$$

And:

$$CEt = \frac{\text{Reduction in Collusion Index}}{\text{Intervention Cost (Time Spent)}}$$

Time spent is currently in minutes, so for consistency, we would convert it to hours:

$$\text{Time Spent (Hours)} = \frac{\text{Time spent (minutes)}}{60 \text{ (hours)}}$$

For both calculations, the reduction in the collusion index would be:

$$\text{Reduction} = CI_{\text{baseline}} - CI_{\text{intervention}}$$

When calculated, the final results are as follows:

Restriction	Cost-effectiveness (CI/CAD)	Time-efficiency (CI/hour)
Memory restriction (30 rounds)	1.4104	0.7990
Memory restriction (0 rounds)	—3.2543 (Ineffective)	—1.1179 (Ineffective)
Financial penalty (25%)	0.6603	0.8661

Competitive disruption

1.2712

0.8948

Table 7. Note: Effectiveness is measured as the reduction in the price-based collusion index relative to the shared baseline. The zero-memory intervention increased the index and is therefore classified as ineffective rather than assigned a positive efficiency score. Cost estimates are allocated Modal expenses, and time-efficiency uses reported experiment runtimes.

Interpreting these results, the 30-round memory restriction, with a cost effectiveness of 1.4104 surpassed all the other restriction groups in terms of modal credits. These findings are preliminary because the Modal costs are estimated allocations and each condition was tested with a single random seed.

On the other hand, the competitive disruption performed the highest in terms of time efficiency with a score of 0.8948. Now comparing the two, the highest performing restriction overall is the memory restriction (Figure 2).

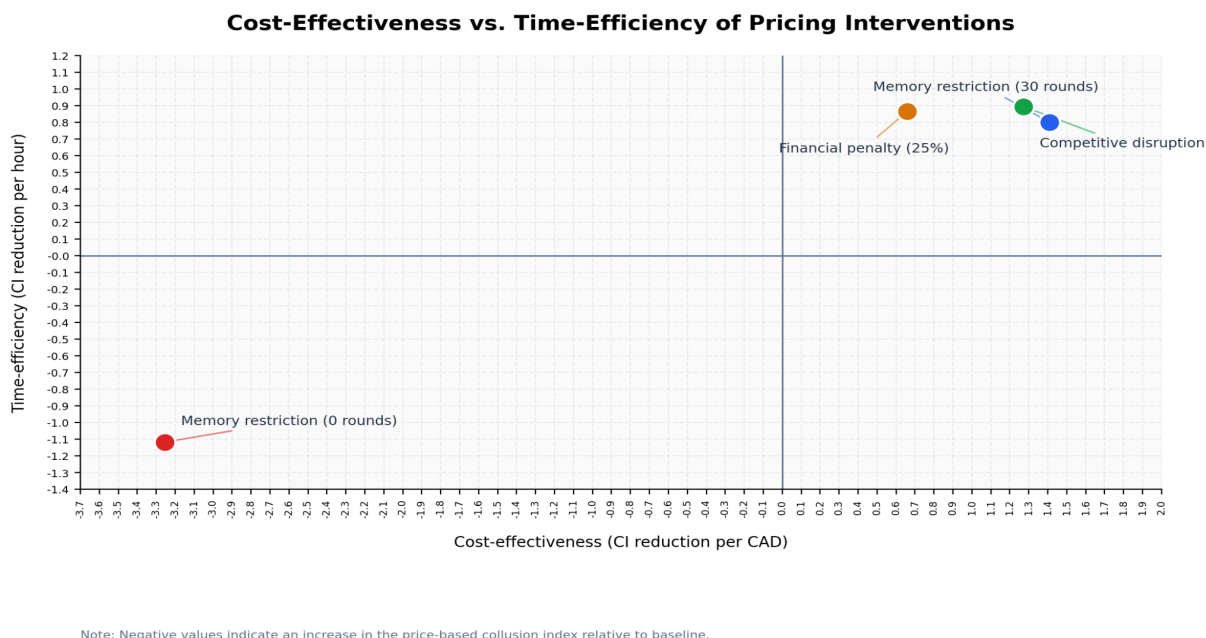


Figure 2

5. Discussion and Limitations

This study has several limitations. First, all experimental conditions were evaluated using a single random seed (0), limiting our ability to assess the consistency of results across independent simulations. Although pricing behavior was observed over 300 rounds, the findings may be sensitive to initial conditions and model variability.

Second, the collusion index measures deviations from competitive pricing rather than directly establishing coordination between AI agents. Prices exceeding the monopoly benchmark, particularly in the baseline and zero-memory conditions, may reflect unstable pricing behavior rather than sustained collusion.

Third, computational costs were estimated using allocated Modal expenditures rather than independently measured costs for each condition. Reported runtimes may also include execution delays, limiting the precision of the cost-effectiveness and time-efficiency comparisons.

Finally, the experiments were conducted using a single LLM architecture within a simulated duopoly. Consequently, the findings may not generalize to other models, market structures, or real-world pricing environments.

6. Conclusion

This study examined the cost-effectiveness of three interventions designed to reduce above-competitive pricing among AI agents: restricting memory access, introducing financial penalties, and competitive disruption. Using a repeated-pricing duopoly, we evaluated each intervention against an unregulated baseline, comparing changes in the price-based collusion index relative to computational expenditure and execution time.

The results indicate that restricting accessible pricing history to 30 rounds achieved the highest estimated cost-effectiveness, producing a reduction of 1.4104 collusion-index units per Canadian dollar of allocated Modal expenditure. Competitive disruption demonstrated the greatest time-efficiency, achieving 0.8948 index units of reduction per hour, while the 25% financial penalty produced the largest absolute reduction in the index. However, removing accessible pricing history entirely increased above-competitive pricing, suggesting that more restrictive interventions do not necessarily produce better outcomes.

Overall, the findings suggest that relatively simple restrictions may offer economically efficient approaches to mitigating elevated AI pricing. Nevertheless, given the exploratory nature of the experiments, further research using multiple random seeds, longer simulation horizons, and additional market environments is necessary to determine whether these results remain consistent and whether the observed pricing patterns reflect sustained algorithmic collusion.

Code and Data

- **Code repository:**
https://github.com/poseidon-research/collusion_hackathon/tree/main/llm-agent-collusion
- <https://github.com/andyjliu/llm-agent-collusion>

Author Contributions

Amanda Okonkwo led the project and designed experiments. Destiny Rock implemented the code. All authors contributed to writing and reviewed the final manuscript.

References

Collina, N., Arunachaleswaran, E. R., & Jagadeesan, M. (2025). Breaking algorithmic collusion in human-AI ecosystems [Preprint]. arXiv. <https://arxiv.org/abs/2511.21935>

Kazemi, Y. (2026). AI collusion Apart hackathon [PowerPoint slides]. Apart Research.

Keppo, J., Li, Y., Tsoukalas, G., & Yuan, N. (2026). On the fragility of AI agent collusion [Preprint]. arXiv. <https://arxiv.org/abs/2603.20281>

Liu, Q., Chen, X., Zhao, Z., Zheng, Y., Yu, D., Zhang, Z., Sun, L., & Chen, Y. (2026). Breaking the secret: Economic interventions for combating collusion in embodied multi-agent systems [Preprint]. arXiv. <https://arxiv.org/abs/2604.23511>

LLM Usage Statement

We used Claude and ChatGPT to help brainstorm approaches and draft sections. All results and claims were independently verified