

MERIT Feedback Elicits Better Bargaining in LLM Negotiators

Jihwan Oh¹ Murad Aghazada¹ Yooju Shin² Se-Young Yun^{1†} Taehyeon Kim^{3†}
¹KAIST AI ²Amazon AGI ³LG AI Research

Abstract

Bargaining is often regarded as a logical arena rather than an art or a matter of intuition, yet Large Language Models (LLMs) still struggle to navigate it due to limited strategic depth and difficulty adapting to complex human factors. Current benchmarks rarely capture this limitation. To bridge this gap, we present a utility feedback centric framework. Our contributions are: (i) AGORABENCH, a new benchmark spanning nine challenging settings (e.g., deception, monopoly) that supports diverse strategy modeling; (ii) human-aligned, economically grounded metrics derived from utility theory. This is operationalized via agent utility, negotiation power, and acquisition ratio that implicitly measure how well the negotiation aligns with human preference and (iii) a human preference grounded dataset with learning pipeline that strengthens LLMs’ bargaining ability through both prompting and finetuning. Empirical results indicate that baseline LLM strategies often diverge from human preferences, while our mechanism substantially improves negotiation performance, yielding deeper strategic behavior and stronger opponent awareness.

1 Introduction

The emergence of Large Language Models (LLMs) as autonomous agents is shifting strategic decision-making and negotiation tasks from humans to LLM-driven systems (Yao et al., 2023; Schick et al., 2023; Park et al., 2023; Xia et al., 2024; Wang et al., 2024; Shea et al., 2024; Bianchi et al., 2024; Abdelnabi et al., 2024; Zhu et al., 2025; Kwon et al., 2025; Chen and Weiss, 2025; Long et al., 2025). Recent work on bilateral negotiation (Deng et al., 2024; Bianchi et al., 2024; Shea et al., 2024; Xia et al., 2024; Kwon et al., 2024; Zhu et al., 2025; Kwon et al., 2025; Chen and Weiss, 2025; Long et al.,

2025) showcases LLMs’ growing proficiency in multi-turn bargaining, underscoring negotiation as a central arena for strategic reasoning. Nonetheless, two core challenges remain: (i) the scarcity of sophisticated benchmarks that capture complex, adversarial bargaining conditions, and (ii) the difficulty of evaluating strategic planning and reasoning, especially in a human-centric sense of what constitutes a *good negotiator* as opponents’ behaviors evolve dynamically beyond simple profit.

(i) Existing benchmarks (He et al., 2018; Lewis et al., 2017; Xia et al., 2024; Fu et al., 2023) largely emphasize simplified, single-issue settings, limiting their usefulness for addressing the weaknesses of current LLM negotiators. Prior works (Xia et al., 2024; Kwon et al., 2024) indicate persistent shortcomings regarding immature Theory of Mind (ToM), constrained strategic adaptability, and often shallow reasoning. These issues are compounded by datasets (He et al., 2018; Lewis et al., 2017; Fu et al., 2023; Xia et al., 2024) that largely ignore common market mechanisms such as installment plans, monopolistic structures, or negative seller perception sentiment. As a result, important facets of reliable bargaining are rarely represented, creating a sizable gap in LLMs’ ability to acquire robust negotiation competencies for intricate, realistic or economically grounded contexts.

To close this gap, we present AGORABENCH, a benchmark that captures diverse, economically motivated market regimes (Table 1). It spans diverse market regimes (e.g., monopolistic, installment-based, negative seller perception) and product types (e.g., cameras, drones), offering a comprehensive testbed for evaluating LLM performance. By incorporating diverse conditions, AGORABENCH provides a strong simulations for probing and improving LLMs’ capacity to manage negotiation intricacies. We also release a human preference dataset derived from LLM-agent dialogues to support in-context learning and fine-tuning.

[†]Corresponding authors.

Code and data are available at <https://github.com/ericoh929/Agorabench.git>.

Features	Bargaining task	Deal or No Deal? (Lewis et al., 2017)	Craiglist (He et al., 2018)	AI feedback (Fu et al., 2023)	AmazonHistoryPrice (Xia et al., 2024)	AGORABENCH (Ours)
Dataset	Economically grounded	✗	✗	✗	✗	✓
	Human Preference	✗	✗	✗	✗	✓
Evaluation	Profit metric	✗	✓	✓	✓	✓
	Acquisition metric	✗	✗	✗	✗	✓
	Human aligned metric	✗	✗	✗	✗	✓

Table 1: Feature comparison between bargaining research. We provide economically grounded market condition and human preference dataset. Further, we devise human preference aligned metric to evaluate the language models.

(ii) Beyond data coverage, steering LLMs toward human-like strategic bargaining requires suitable evaluation and feedback. Since LLMs may otherwise default to naive optimization, we introduce MERIT, a new, multi-dimensional metric designed to align behavior with human preferences, moving past conventional profit-centric assessments (He et al., 2018; Lewis et al., 2017; Xia et al., 2024; Fu et al., 2023). We observe that profit-alone measures miss genuine human priorities (§4.2). For example, in multi-product settings, securing the desired item can be valued more than minimizing price. By combining profit with acquisition ratio, MERIT better captures human preferences by accounting for both the negotiation process and its outcomes. In our benchmark, higher MERIT indicates a negotiator that behaves more consistently with human bargaining preferences, rather than optimizing numerical utility alone.

To this end, we utilize MERIT as a feedback to promote exploration of strategic planning and reasoning against adaptive opponents. Integrating this metric into negotiation via an in-context learning (ICL) and fine-tuning enables LLMs to internalize the nuanced trade-offs of realistic bargaining. Table 1 highlights the greater scenario diversity and human-aligned evaluation of our framework relative to prior work. Our contributions are as follows.

- AGORABENCH. A benchmark capturing a wide range of scenarios that reflect realistic market environments. (§3)
- MERIT. A novel, multi-faceted evaluation metric that comprehensively assesses buyer-side bargaining performance. Grounded in economic theory, it aligns with human preferences over both process and outcomes. (§4)
- We identify negotiation behaviors exhibited by LLM negotiators that deviate from typical human negotiating patterns. (§5)
- We show that both MERIT guided ICL and finetuning advances LLM strategic negotia-

tion, raising deal rates over multiple models (§6).

2 Related works

2.1 Negotiation and bargaining tasks

Negotiation has drawn growing interest within natural language processing as language models increasingly act as autonomous agents capable of goal-directed, multi-turn dialogue (Kwon et al., 2024; Bianchi et al., 2024; Guan et al., 2024; Hua et al., 2024; Shea et al., 2024; Zhu et al., 2025; Kwon et al., 2025). A foundational line of bargaining work is Lewis et al. (2017), which introduced the *Deal or No Deal?* human-human bargaining corpus and examined end-to-end models for negotiating item splits. Building on this, research shifted toward more structured strategy generation. For instance, He et al. (2018) proposed a modular approach that separates high-level strategy planning from surface realization and released the *Craiglist* bargaining dataset. Fu et al. (2023) designed a simplified balloon-product bargaining task and leveraged LLM critiques as feedback for in-context learning. Xia et al. (2024) constructed a large-scale bargaining dataset by harvesting Amazon product data from the camelcamelcamel platform. However, as summarized in Table 1, these resources still fall short of the diversity and real-world complexity required for robust evaluation.

2.2 Model alignment with human preferences

As language models increasingly operate as autonomous decision-makers with strategic reasoning capabilities, aligning their behavior with human preferences has become a central research objective. In negotiation and bargaining, performance must be judged not only by quantitative scores but also by subtler, human-centric priorities—such as fairness, perceived satisfaction, and trustworthiness (Weidinger et al., 2021; Bai et al., 2022; Perez et al., 2023).

A substantial body of work on reward alignment seeks reward designs that more faithfully capture human intent, moving beyond coarse metrics like

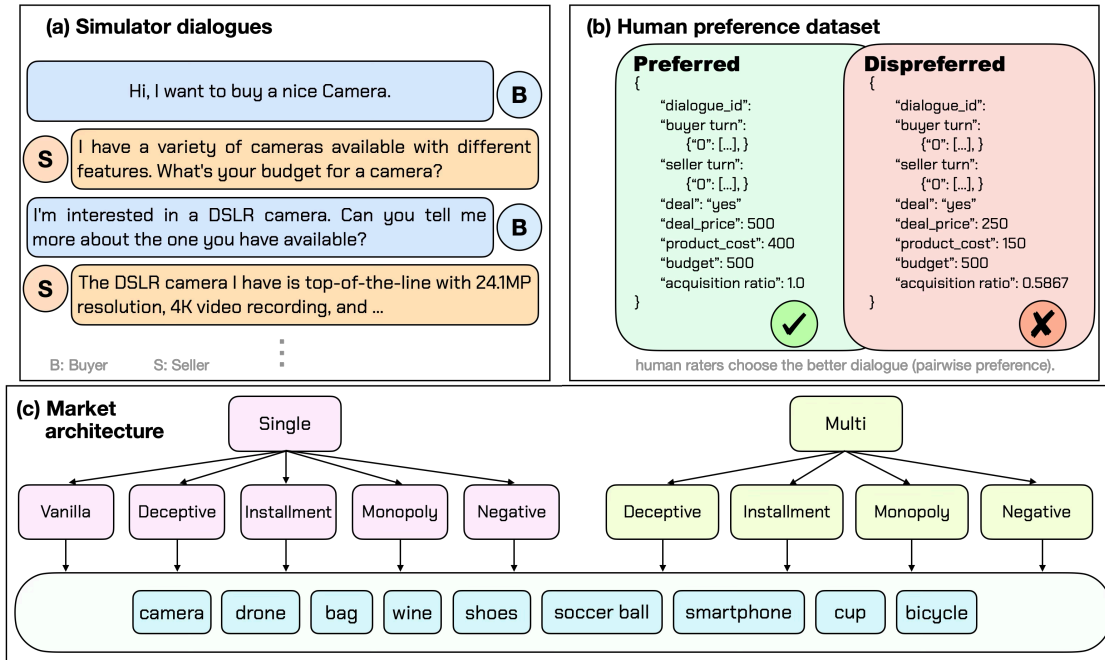


Figure 1: Negotiation task between LLM agents and AGORABENCH overview. (a) represents the simulator, (b) represents the human preference dataset, and (c) represents nine economically grounded market environments, each crafted around a distinct consumer good.

task accuracy or profit alone. Preference-based approaches, for example, leverage human comparisons and feedback to steer reinforcement learning agents toward human values (Christiano et al., 2017; Knox et al., 2022; Ouyang et al., 2022). Related efforts introduce multi-objective reward formulations to encode trade-offs among competing goals—e.g., short-term profit versus long-term customer satisfaction (Hwang et al., 2024; Wang et al., 2024). Despite these advances, comparatively little attention has been paid to embedding human preferences directly into the evaluation metrics used for bargaining tasks.

3 AGORABENCH

AGORABENCH mirrors real-world economic bargaining across nine distinct market types, curated to surface specific commercial negotiation challenges. It provides both an online simulator (§3.1) and a static offline dataset (§3.2). To ensure negotiation is necessary, we assume the seller’s initial ask always exceeds the buyer’s willingness to pay (WTP). We adopt the term *WTP* over *budget* to more precisely reflect economic preferences.

3.1 Simulator environment

Two LLM agents engage in negotiations within the simulator using a structured *Thoughts–Talk–Action* schema (Yao et al., 2023; Xia et al., 2024), also referred to as ReAct (Figure 1 (a)). Each agent

operates within a constrained action space (BUY, SELL, REJECT, DEAL, and QUIT) following the framework established by Xia et al. (2024). Each dialogue is grounded in a specific buyer–seller scenario defined by key hyperparameters: the market setting, agent profiles, and product attributes such as cost, willingness to pay (WTP), and initial ask. The AGORABENCH simulator encompasses nine distinct market regimes, constructed by instantiating five market settings under the SINGLE condition and four additional settings under the MULTI condition (Figure 1 (c)). A full specification of the dataset format and fields is provided in Appendix A.

- **Vanilla market:** A baseline, product negotiation with no additional complicating factors.
- **Deceptive market:** Settings in which either party may misrepresent information (e.g., product quality), testing agents’ ability to identify reliable signals. Except for the vanilla case, all other markets are built on this deception layer, with further conditions added atop it.
- **Monopoly market:** A single-seller environment that limits buyer choice, enabling analysis of bargaining under asymmetric power.
- **Installment possible market:** Introduces deferred or staggered payments, bringing time-sensitive financial trade-offs and flexible terms into the dialogue.

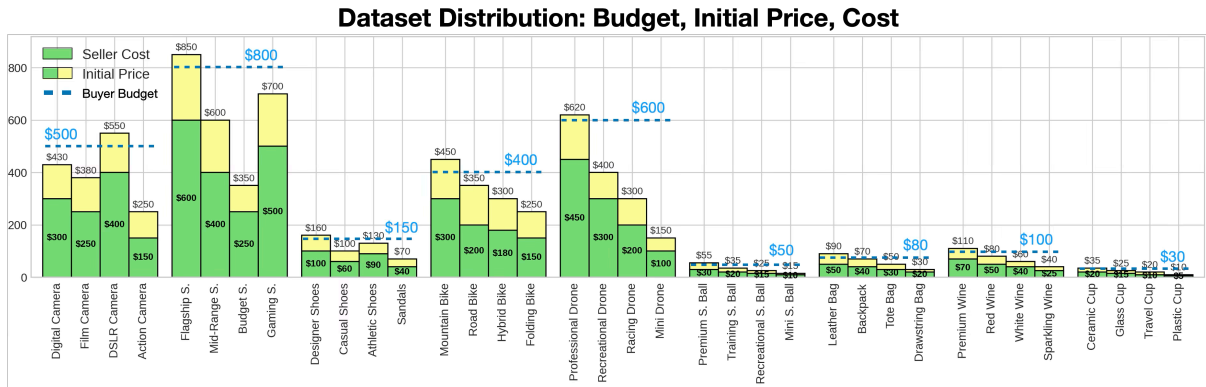


Figure 2: Product attributes for negotiation scenarios in *multi* product setting. We deploy seller cost, initial offer price, and buyer’s budget for each product.

- **Negative perception market:** The seller carries a prior reputational disadvantage (e.g., scandals), biasing buyers toward lower offers and creating distinct strategic dynamics as the seller attempts to counteract this effect.
- **Single product market:** Buyers negotiate over a single item with no substitution across alternatives. This isolates bargaining behavior in a focused setting where outcomes depend solely on the agreed price and the buyer’s willingness-to-pay.
- **Multi product market:** Buyers consider multiple items and can substitute away from a preferred option due to constraints (e.g., budget). This highlights the trade-off between product preference and cost savings. We layer this condition atop each non-vanilla market.

3.2 Human preference

We further present a human preference dataset for negotiation over multiple LLM negotiations. Precisely, negotiation dialogues by multiple LLMs in deceptive multi-product negotiation settings across all set of products are collected, and then human preferences are obtained via surveys on Amazon Mechanical Turk (MTurk)¹. Data is annotated in a standardized JSONL format as shown in Figure 1 (b). Using this dataset, we demonstrate how human preference signals can be leveraged to improve negotiation performance in section 6.

4 Human-Aligned Utility Metrics

We introduce MERIT, a Multi-dimensional Evaluation of Reasoning & Interaction in Trade, grounded in the economic principles of *utility* and *bargaining*. In standard microeconomics (Mankiw

¹<https://www.mturk.com/>

et al., 2007; Jehle, 2011), rational agents seek to maximize *utility*, a construct that encompasses more than pure profit. In line with this perspective, MERIT evaluates not only monetary gains but also an agent’s overall satisfaction from the exchange (total utility) and the degree to which the chosen product aligns with prior *preferences* (ordinal utility). Thus, MERIT captures both the *cardinal* and *ordinal* facets of utility that are essential for realistic evaluation of bargaining tasks.

4.1 MERIT: Formulation

The performance of a language model as a buyer is quantified by MERIT through a score reflecting how effectively the agent achieves its objectives within a given negotiation scenario. In microeconomic terms, this score corresponds to the *overall satisfaction* or *total benefit* an agent derives from a transaction, paralleling the notion of *cardinal utility* where a measurable value is assigned to outcomes. Specifically, the MERIT score for a buyer is a weighted combination of three key components: *Consumer Surplus (CS)*, *Negotiation Power (NP)*, and *Acquisition Ratio (AR)*. Each component captures a distinct economic concept related to consumer behavior, bargaining dynamics, and preference fulfillment, as defined in Equation 1:

$$\text{MERIT}_{\text{buyer}} = \alpha \times \text{CS} + \beta \times \text{NP} + \gamma \times \text{AR} \quad (1)$$

Consumer surplus In economics, *consumer surplus* refers to the net benefit (or satisfaction) a consumer derives from purchasing a product or service (Mankiw et al., 2007). It traditionally measures the difference between the consumer’s *willingness to pay* (P_{wtp}) and the actual price paid (P_{deal}). Within MERIT, we normalize this surplus relative to the potential surplus range determined by the seller’s production cost (P_{cost}):

$$\text{Consumer Surplus (CS)} = \frac{P_{\text{wtp}} - P_{\text{deal}}}{P_{\text{wtp}} - P_{\text{cost}}} \quad (2)$$

Bounding CS between 0 and 1 translates consumer surplus into a scaled, *cardinal measure* of how effectively the buyer maximized net benefits. A value approaching 1 indicates a highly favorable deal, while 0 suggests an unfavorable outcome or an infeasible price.

Negotiation power Bargaining outcomes also hinge on the buyer’s ability to *shift* the final price in their favor—an aspect tied to the broader economic concept of *bargaining power* (Chatterjee and Samuelson, 1983; Muthoo, 1999). MERIT incorporates this through:

$$\text{Negotiation Power (NP)} = \frac{P_{\text{initial}} - P_{\text{deal}}}{P_{\text{initial}} - P_{\text{cost}}} \quad (3)$$

where P_{initial} is the seller’s initial asking price. A higher NP indicates that the buyer successfully negotiated the deal price downward from the initial offer. In classic bargaining theory (e.g., Nash bargaining (Binmore et al., 1986)), each party’s utility share depends on their relative power. Here, NP complements the *cardinal utility* of consumer surplus by capturing how effectively the buyer reclaims surplus that might otherwise go to the seller. Values near 1 imply maximal movement from the seller’s initial price, reflecting strong buyer influence, whereas values near 0 imply that the buyer gained little ground in negotiations.

Acquisition ratio of desired goods Finally, a consumer’s satisfaction also depends on obtaining the *right* product, reflecting the ordinal aspect of utility—that is, how closely the outcome aligns with the consumer’s preference ranking. We quantify this via an *Acquisition Ratio*, which measures how semantically similar the buyer’s acquired items are to their desired items. By embedding items in a high-dimensional space using text-embedding-ada-002, we compute:

$$\text{Acquisition Ratio (AR)} = \frac{\mathbf{v}_{\text{acquired}} \cdot \mathbf{v}_{\text{desired}}}{\|\mathbf{v}_{\text{acquired}}\| \|\mathbf{v}_{\text{desired}}\|} \quad (4)$$

where $\mathbf{v}_{\text{acquired}}$ and $\mathbf{v}_{\text{desired}}$ represent embedding vectors of the items actually purchased and initially sought, respectively. A higher AR indicates that the buyer’s outcome closely matches their *top preferences*, embodying the idea that utility is also determined by which specific goods are consumed, not merely how profitable the purchase is.

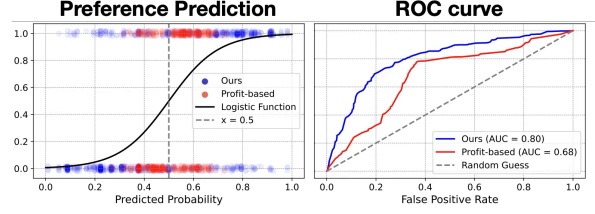


Figure 3: Demonstration that a human-aligned metric outperforms a purely profit-based one.

4.2 Human preference alignment

To ensure MERIT reflects human judgments, we ran an Amazon Mechanical Turk (MTurk) study. Participants were shown pairs of negotiation dialogues (20 comparisons per product; 180 participants total) and asked to choose the better one. The survey included both same-item price comparisons and trade-offs between similar items; full details are in Appendix C.

The coefficients α, β, γ in Equation 1 were optimized globally using merged data from all product categories based on Bradley-Terry model (Bradley and Terry, 1952). Since only relative magnitudes matter in the Bradley-Terry model, we scale them for interpretability so that their sum is 3, as follows:

$$\alpha' = \frac{\alpha}{\alpha + \beta + \gamma} \times 3, \quad \beta' = \frac{\beta}{\alpha + \beta + \gamma} \times 3, \quad \gamma' = \frac{\gamma}{\alpha + \beta + \gamma} \times 3.$$

After optimization, the final scaled coefficients are: $\alpha' = 1.0139, \beta' = 0.8812, \gamma' = 1.1049$.

As shown in Figure 3, this process confirmed that MERIT aligns more closely with human preferences than traditional profit-based ones (e.g., achieving an ROC AUC of 0.80 vs. 0.68 for the latter). This human-aligned metric is subsequently used for model evaluation and as feedback in our system. Details on the optimization of the metric model are provided in Appendix C.

5 Behaviors of LLM Negotiators

In this section, we evaluate the negotiation capabilities of contemporary LLMs, specifically the GPT-series and Gemini-series, within our AGORABENCH. The experiments are designed to uncover emergent behaviors and provide insights into LLM performance across varied strategic contexts.

Anchoring effect and first-mover advantage (Table 3) We investigated the *anchoring effect* (Orr and Guthrie, 2005), a cognitive bias where initial information heavily influences subsequent decisions. Experiments in the vanilla market, using two distinct initial camera prices (\$550 and \$520,

Models	Vanilla		Deceptive		Monopoly		Installment		Negative Perc.		Avg
	Single	Single	Multi	Single	Multi	Single	Multi	Single	Multi		
GPT-series											
gpt-4o	1.414 ± 0.07 (86.7%)	1.342 ± 0.09 (86.4%)	1.275 ± 0.16 (84.6%)	1.139 ± 0.11 (78.9%)	1.211 ± 0.13 (84.6%)	1.188 ± 0.09 (84.4%)	1.032 ± 0.17 (78.6%)	1.281 ± 0.17 (83.9%)	1.342 ± 0.11 (87.4%)	1.247 (83.9%)	
gpt-4o-mini	1.164 ± 0.11 (77.1%)	1.228 ± 0.12 (74.4%)	<u>1.402</u> ± 0.11 (89.7%)	<u>1.230</u> ± 0.14 (79.6%)	1.381 ± 0.11 (89.9%)	1.105 ± 0.16 (78.9%)	1.376 ± 0.19 (88.5%)	0.973 ± 0.16 (56.0%)	1.305 ± 0.17 (76.7%)	1.240 (79.0%)	
gpt-3.5-turbo	<u>1.322</u> ± 0.11 (94.4%)	<u>1.366</u> ± 0.12 (90.0%)	1.344 ± 0.14 (90.5%)	1.159 ± 0.07 (81.3%)	1.159 ± 0.07 (95.8%)	<u>1.254</u> ± 0.08 (90.7%)	1.177 ± 0.12 (84.6%)	1.140 ± 0.12 (78.2%)	<u>1.395</u> ± 0.18 (91.0%)	<u>1.257</u> (88.5%)	
Gemini-series											
gemini-1.5-pro	1.202 ± 0.15 (72.0%)	1.394 ± 0.19 (75.8%)	1.502 ± 0.09 (93.6%)	1.350 ± 0.19 (74.6%)	<u>1.324</u> ± 0.13 (66.7%)	1.522 ± 0.09 (89.4%)	<u>1.308</u> ± 0.13 (94.2%)	1.300 ± 0.40 (68.0%)	1.525 ± 0.18 (88.8%)	1.381 (80.3%)	
gemini-1.5-flash	0.875 ± 0.23 (61.8%)	1.005 ± 0.21 (66.3%)	1.304 ± 0.15 (87.6%)	0.951 ± 0.17 (61.8%)	1.110 ± 0.20 (65.0%)	0.988 ± 0.20 (71.8%)	1.056 ± 0.10 (75.4%)	0.826 ± 0.18 (54.4%)	1.284 ± 0.17 (82.6%)	1.044 (69.6%)	

Table 2: Buyer MERIT performance with standard deviation using the ReAct (Yao et al., 2023) framework across different market conditions. Each cell presents the MERIT score (top) and deal rate (bottom, in parentheses). **Bold** indicates the best performance, and underline indicates the second best.

Buyers	Market 1*		Market 2**		Market 3***	
	AVG price ↓	MERIT ↑	AVG price ↓	MERIT ↑	AVG price ↓	MERIT ↑
gemini-1.5-pro	444.6	2.18	432.1	2.38	422.6	2.39
gemini-1.0-pro	470.2	1.63	455.4	2.05	434.7	2.11
gemini-1.5-flash	481.4	1.62	476.0	1.68	485.6	1.61

*: initial price 550 **: initial price 520 ***: Buyer propose the price first.

Table 3: Anchoring effect and first-mover advantage.

denoted Market 1* and Market 2** respectively) set by the seller, demonstrated this bias: higher initial anchor prices generally led to higher final deal prices. Furthermore, when the buyer initiated the first offer (Market 3***), the average deal price was even lower compared to scenarios where the seller proposed first, highlighting a potential first-mover advantage for the buyer in price setting.

Irrational concessions: Unstable anchoring (Table 20) We observed unnatural negotiation behavior in LLMs, especially smaller models such as gpt-4o-mini. In Table 20, the buyer shows unstable anchoring: after proposing a price, it reverses direction with strictly lower counteroffers, effectively backtracking on its stated stance. Concretely, the buyer drops from \$480 to \$475 and \$450, then later returns to \$480 before agreement. Such reversed concessions are atypical in human bargaining, where offers are generally monotone toward settlement, and they weaken the buyer’s anchor by signaling inconsistent *willingness to pay* reflecting anchor instability rather than strategic bargaining.

Impact of market conditions (Table 2)

- **Deceptive.** Allowing buyers to use deception generally *improves buyer outcomes*: MERIT increases vs. vanilla (avg. +0.10) and deal rates rise substantially (+10.62%p). Some GPT models trade slightly lower deal rates for higher scores (e.g., gpt-4o), whereas Gemini often improves both, suggesting a clearer buyer advantage.

- **Monopoly.** Monopoly consistently harms buyers across models: deal rates drop by 3.34% (single-item) and 8.80% (multi item), while MERIT declines by about 0.10 and 0.13, respectively. This reflects increased seller leverage, higher prices, and weaker convergence.
- **Installment.** Installment options have mixed effects: in single-item settings they often increase deal rates but at the cost of worse buyer prices (lower MERIT); in multi-item settings, added bargaining complexity tends to reduce deal rates overall.
- **Negative perception.** A negative prior about the seller consistently reduces deal rates, with a sharper effect in single-item negotiations (few outside options). In multi-item settings, substitutes provide limited mitigation but do not eliminate the decline.
- **Multi product.** Across conditions, multi-item settings generally exhibit higher deal rates than single-item ones, consistent with the presence of substitutes that facilitate agreement. However, this benefit is not uniform: under *installment*, GPT models show a reversal where deal rates drop despite substitutes. Such sign reversals serve as a useful diagnostic of market compatibility, indicating failures to leverage multi product option when bargaining becomes multi-dimensional.

By comparing each market condition against the benchmark’s baseline (*deceptive* or vanilla) in terms of both deal rate and buyer MERIT, we can assess whether LLM’s *market compatibility* they align with patterns typical in human bargaining or deviates from them. In particular, directional shifts across conditions (e.g., reduced agreement

Model (vs 1.5-pro)	Single product			Multi product		
	ReAct	OG-Narrator	Ours	ReAct	OG-Narrator	Ours
GPT-4o	1.127 84.2%	1.094 51.6%	1.662 100%	1.045 76.1%	1.478 82.8%	1.563 99.7%
GPT-4o-mini	1.143 73.0%	1.165 45.1%	1.450 97.1%	1.493 96.4%	1.281 66.1%	1.527 98.6%
GPT-3.5-turbo	1.248 86.9%	1.157 45.5%	1.510 99.1%	1.119 82.8%	1.346 71.9%	1.456 100%
Gemini-1.5-pro	1.354 76.0%	1.197 43.8%	1.841 99.8%	1.638 97.2%	1.544 81.9%	1.768 100%
Gemini-1.5-flash	0.929 63.2%	0.689 26.9%	1.418 99.8%	1.506 98.9%	1.422 80.3%	1.565 99.7%
Average	1.160 76.7%	1.060 42.6%	1.576 99.2%	1.360 90.3%	1.414 76.6%	1.576 99.6%
Model (vs DeepSeek-v3.2)						
GPT-5.1	1.849 99.6%	1.423 50.4%	2.027 99.1%	1.950 100%	1.536 85.8%	2.095 99.7%
DeepSeek-v3.2	1.675 100%	0.333 12.2%	1.875 98.9%	1.628 100%	0.957 52.5%	1.872 100%
Sonnet 4.5	1.457 99.8%	1.856 67.1%	1.778 99.8%	1.560 100%	1.368 53.1%	1.890 100%
Gemini-3-Flash	1.870 100%	2.116 76.7%	1.995 99.8%	1.856 100%	1.346 56.4%	2.035 99.7%
gpt-oss-120b	1.318 89.6%	0.098 4.9%	1.391 90.6%	1.268 86.4%	0.934 60.3%	1.431 90.3%
gpt-oss-20b	1.325 89.3%	0.154 5.6%	1.413 91.3%	0.986 65.3%	0.545 33.8%	1.410 86.6%
Average	1.582 96.4%	0.997 36.2%	1.747 96.6%	1.541 92.0%	1.114 57.0%	1.789 96.1%
Model (vs DeepSeek-v3.2)						
GPT-5.1 reasoning mode	1.999 100%	1.315 46.9%	2.127 100%	2.055 100%	1.452 78.9%	2.189 99.7%
DeepSeek-v3.2 reasoning mode	1.648 99.8%	0.275 10.4%	1.852 98.9%	1.632 99.2%	0.900 49.2%	1.885 99.4%
Average	1.824 99.9%	0.795 28.7%	1.990 99.5%	1.844 99.6%	1.176 64.1%	2.037 99.6%

Table 4: We compare ICL-MF with ReAct and OG-Narrator using MERIT and deal rate. We run 10 dialogues per product under each market condition, and report averages separately for the single-item and multi-item settings. Cells are highlighted in pink when ICL-MF achieves the best performance, and in light purple when OG-Narrator achieves the best performance.

and utility under *monopoly*, or trust-driven drops under *negative perception* reveal whether a model responds to changes in leverage, multiple options, and priors in an economically coherent manner. For detailed explanation of LLMs behaviors, see in Appendix E.

6 MERIT-Guided Prompting & Training

To address the common misalignment of LLM-based negotiations with human preferences, we propose MERIT-guided In-Context Learning (ICL-MF) and training to enhance negotiation strategies.

6.1 Human preference evaluation (MERIT)

MERIT-guided Prompting over Others. We compared ICL-MF with two baselines; ReAct (Yao et al., 2023), which prompts LLMs to reason before acting, and OG-Narrator (Xia et al., 2024), which generates offers by linearly interpolating the buyer’s price from 50% to 100% of the budget across turns. We evaluate the performance with MERIT for five buyer (closed-sourced) LLMs (GPT/Gemini series) negotiating against a gemini-1.5-pro seller. Further, to assess scalability across model families, we evaluate six buyer LLMs—GPT-5.1 (Singh et al., 2025), DeepSeek-v3.2 (Liu et al., 2025), Sonnet 4.5 (An-

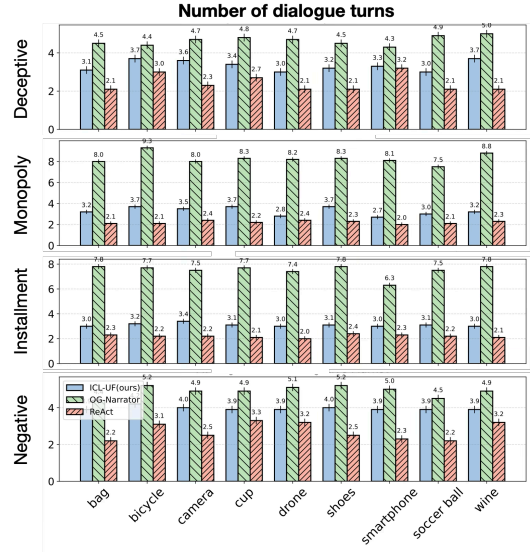


Figure 4: Number of turns for the negotiation. We feed dialogues from gpt-4o buyer and gemini-1.5-pro seller to gemma-3-27b judge.

thropic, 2025), Gemini-3-Flash (Google DeepMind, 2025) (closed-source) and gpt-oss-120b, gpt-oss-20b (open-source; the gpt-oss series (Agarwal et al., 2025))—against a DeepSeek-v3.2 seller. Also we conduct experiments based on the reasoning model (switching on the reasoning mode of GPT-5.1 and DeepSeek-v3.2) to observe whether performance gain exist in the reasoning mode. As shown in Table 4, ICL-MF consistently outperforms both baselines by a large margin in terms of MERIT and deal rate in both single- and multi-product settings, and these gains hold consistently across closed-sourced models, open-sourced models, and reasoning variants.

Analysis of results. OG-Narrator often underperforms ReAct due to its rigid price schedule, which fails to adapt to opponent temperament. Against patient sellers like gemini-1.5-pro, it excessively prolongs dialogue (Figure 4), leading to bargaining impasses (Babcock and Loewenstein, 1997; Hwang, 2018; Backus et al., 2020), without reaching a deal. Conversely, with impatient sellers like deepseek-v3.2, negotiations terminate abruptly as opponents exit prematurely. While ReAct settles too quickly due to premature assumptions (Laban et al., 2025), ICL-MF balances efficiency and effectiveness through appropriate negotiation lengths. Crucially, ICL-MF demonstrates a clear positive correlation between model capability and performance, better leveraging the inherent strengths of stronger LLMs.

We additionally report raw economic outcomes beyond MERIT: *Deal Rate*, *Average Deal Price*

Method	Exemplary Thought Process
ReAct	Thought1: "\$475 is still a bit high. I'll push a little further and feign disinterest to see if they'll budge."
	Thought2: "\$550 is above my budget. I'll start low and express hesitation... I'll hint I might walk away to add pressure."
ICL-MF	Thought1: "They moved a decent amount, suggesting their cost is probably closer to \$400. If their cost is \$400, a deal at \$450 would give me $CS = 0.5$ and $NP = 0.67$."
	Thought2: "I'm guessing they might be willing to go a bit lower—maybe their cost is closer to \$350. At \$450, CS would be 0.33 and NP would be 0.5. I need to make them think I might walk away."

Table 5: Comparison of internal reasoning: ReAct vs. ICL-MF with gemini-1.5-pro. While ReAct uses qualitative tactics, ICL-MF quantifies utility through explicit cost estimation and economic metrics (CS , NP).

gpt-oss-20b (vs DeepSeek-v3.2)	Single product		Multi product	
	MERIT	Deal rate	MERIT	Deal rate
ReAct	1.325	89.3%	0.986	65.3%
ICL-MF (ours)	1.413	91.3%	1.410	86.6%
SFT (ours)	<u>1.332</u>	<u>89.6%</u>	1.459	93.1%

Table 6: We compare ReAct with our data-utilized methods; ICL and SFT utilizing gpt-oss-20b. **Bold** indicates the best performance, and underline indicates the second best.

(deals only), and Average Buyer Profit. Average Deal Price is computed over successful negotiations only and should be interpreted jointly with Deal Rate. As shown in Table 19 (Appendix D.2), these complementary metrics are consistent with our main results, alleviating the single evaluation pathway concern.

MERIT elicits opponent-aware reasoning.

Adapting MERIT as guidance for negotiation encourages LLMs to transition from simple tactic-centric thoughts to deep opponent-aware reasoning (OAR). As illustrated in Table 5, while the ReAct baseline relies on vague behavioral maneuvers such as “feigning disinterest” or “adding pressure”, ICL-MF enables the agent to explicitly hypothesize the seller’s hidden belief, specifically their underlying cost. For instance, the ICL-MF agent deduces that the opponent’s cost is likely “closer to \$400” based on their previous moves, subsequently calculating its own Consumer Surplus (CS) and Negotiation Power (NP) to evaluate the feasibility of a deal.

Human preference dataset training We enhance the negotiation power of LLMs by fine-tuning gpt-oss-20b using LoRA on human-preferred dialogues (Figure 1 (b)). To maintain practical realism, we exclude unobservable seller thoughts from the training data. We define a filtered preferred dialogue as a turn sequence $\mathcal{D} = \{(b_t^r, b_t^u, b_t^a, s_t^u, s_t^a)\}_{t=0}^{T-1}$, where b_t^r , b_t^u , and b_t^a denote the buyer’s reasoning, utterance, and action at

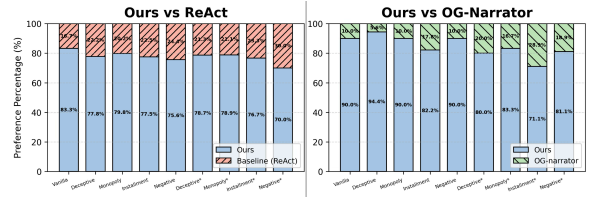


Figure 5: Preference comparison between ICL-MF vs ReAct; OG-Narrator algorithm by LLM-judge

turn t , respectively. The model is trained to predict the buyer’s next turn y_t given the interaction history x_t :

$$\begin{aligned}
 b_t &= (b_t^r, b_t^u, b_t^a), & s_t &= (s_t^u, s_t^a) \\
 x_t &= [b_0, s_0, b_1, s_1, \dots, s_{t-1}], & (5) \\
 y_t &= b_t, & t &= 0, 1, \dots, T-1
 \end{aligned}$$

The buyer policy π_θ is optimized using the standard maximum-likelihood objective:

$$\mathcal{L}_{\text{SFT}}(\theta) = - \sum_{\mathcal{D}} \sum_{t=0}^{T-1} \log \pi_\theta(y_t | x_t) \quad (6)$$

As shown in Table 6, the SFT gpt-oss-20b consistently outperforms the ReAct baseline when simulated against deepseek-v3.2. Notably, while ICL-MF excels in single-product settings due to efficient situational matching, SFT proves superior in multi-product negotiations that require stable management of diverse products and complex trade-offs. These results demonstrate that leveraging human negotiation strategies leads to more robust outcomes than pure reasoning-based approaches.

Importantly, the SFT model is trained only in the Deceptive regime with GPT/Gemini-family opponents but evaluated across all nine regimes and against an unseen DeepSeek-v3.2 seller, constituting a dual out-of-distribution shift.

6.2 LLM preference evaluation

We also assessed generated negotiation dialogues using an LLM as an impartial judge. This aligns with the emerging paradigm of leveraging LLMs

for fine-grained evaluation (Zheng et al., 2023; Chiang et al., 2023), offering a scalable means to capture interaction aspects that automated scores might overlook. For this task, gemma-3-27b-it (Team, 2025) evaluated dialogues from ICL-MF (gpt-4o buyer vs. gemini-1.5-pro seller) against those from two baselines: ReAct and OG-Narrator. The LLM-judge was presented with paired dialogues in random order and prompted to select the negotiation more favorable to the buyer, considering outcome, budget adherence, and product features (see Appendix B for the detailed prompt).

The results in Figure 5 show a clear preference for ICL-MF across all market conditions. Against OG-Narrator, ICL-MF often exceeded 84.7% preference on average, peaking at 94.4% in the deceptive market and remaining strong (e.g., 71.3%) even under installment scenarios. Against ReAct, ICL-MF likewise outperformed with preference rates typically between 70% and over 83.3%. Overall, this independent LLM-judge evaluation corroborates our main findings: ICL-MF not only optimizes the targeted utility components but also produces buyer-perceived strategies and outcomes that are more favorable and strategically sound.

7 Conclusion

We introduce a dataset and benchmark that mirror economically grounded market complexity, a human-aligned multi-facet metric, and an in-context learning and training method that uses this metric as feedback to enhance negotiation skills. By capturing both profit and acquisition priorities, our framework fills gaps left by profit-only metrics and, we hope, helps the community build more effective, human-aligned LLM bargaining agents.

Limitations

While our work advances LLM bargaining, several limitations suggest avenues for future exploration. First, our study is primarily buyer-centric, overlooking seller-specific objectives such as inventory risk or tiered profit targets; future work should incorporate a dedicated seller perspective. Second, while AGORABENCH covers diverse scenarios, it does not yet capture global market subtleties like regulatory contexts or cultural norms. Third, we do not cover tool-augmented agentic environments (e.g., LLM-agent tool use for local retrieval or search), which could materially affect bargaining behavior by enabling grounded, real-time access to exter-

nal information (e.g., comparable prices or product specifications). While building such environments would be valuable, it is beyond the scope of this study; we leave tool-integrated settings (including local search) to future work.

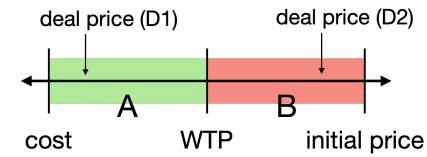
References

- Sahar Abdelnabi, Amr Gomaa, Sarath Sivaprasad, Lea Schönherr, and Mario Fritz. 2024. Cooperation, competition, and maliciousness: Llm-stakeholders interactive negotiation. *Advances in Neural Information Processing Systems*, 37:83548–83599.
- Sandhini Agarwal, Lama Ahmad, Jason Ai, Sam Altman, Andy Applebaum, Edwin Arbus, Rahul K Arora, Yu Bai, Bowen Baker, Haiming Bao, and 1 others. 2025. gpt-oss-120b & gpt-oss-20b model card. *arXiv preprint arXiv:2508.10925*.
- Anthropic. 2025. [Claude sonnet 4.5 system card](#).
- Linda Babcock and George Loewenstein. 1997. Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic perspectives*, 11(1):109–126.
- Matthew Backus, Thomas Blake, Jett Pettus, and Steven Tadelis. 2020. Communication and bargaining breakdown: An empirical analysis. Technical report, National Bureau of Economic Research.
- Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, and 1 others. 2022. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*.
- Federico Bianchi, Patrick John Chia, Mert Yuksekgonul, Jacopo Tagliabue, Dan Jurafsky, and James Zou. 2024. How well can llms negotiate? negotiationarena platform and analysis. *arXiv preprint arXiv:2402.05863*.
- Ken Binmore, Ariel Rubinstein, and Asher Wolinsky. 1986. The nash bargaining solution in economic modelling. *The RAND Journal of Economics*, pages 176–188.
- Ralph Allan Bradley and Milton E. Terry. 1952. [Rank analysis of incomplete block designs: I. the method of paired comparisons](#). *Biometrika*, 39(3/4):324–345.
- Kalyan Chatterjee and William Samuelson. 1983. Bargaining under incomplete information. *Operations research*, 31(5):835–851.
- Siqi Chen and Gerhard Weiss. 2025. Enhancing agent-based negotiation strategies via transfer learning. *Electronics*, 14(17):3391.

- Wei-Lin Chiang, Zhuohan Li, Ziqing Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, and 1 others. 2023. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. See <https://vicuna.lmsys.org> (accessed 14 April 2023), 2(3):6.
- Younwoo Choi, Changling Li, Yongjin Yang, and Zhi-jing Jin. 2025. Agent-to-agent theory of mind: Testing interlocutor awareness among large language models. *arXiv preprint arXiv:2506.22957*.
- Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30.
- Yuan Deng, Vahab Mirrokni, Renato Paes Leme, Hanrui Zhang, and Song Zuo. 2024. Llms at the bargaining table. In *Agentic Markets Workshop at ICML*, volume 2024.
- Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. Improving language model negotiation with self-play and in-context learning from ai feedback. *arXiv preprint arXiv:2305.10142*.
- Google DeepMind. 2025. [Gemini 3 flash model card](#).
- Zhenyu Guan, Xiangyu Kong, Fangwei Zhong, and Yizhou Wang. 2024. Richelieu: Self-evolving llm-based agents for ai diplomacy. *Advances in Neural Information Processing Systems*, 37:123471–123497.
- He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. *arXiv preprint arXiv:1808.09637*.
- Kaitlyn Henderson. 2024. The crisis of low wages: Who earns less than \$17 an hour in the us in 2024?
- Wenyue Hua, Ollie Liu, Lingyao Li, Alfonso Amayuelas, Julie Chen, Lucas Jiang, Mingyu Jin, Lizhou Fan, Fei Sun, William Wang, and 1 others. 2024. Game-theoretic llm: Agent workflow for negotiation games. *arXiv preprint arXiv:2411.05990*.
- Ilwoo Hwang. 2018. A theory of bargaining deadlock. *Games and Economic Behavior*, 109:501–522.
- Minyoung Hwang, Luca Weihs, Chanwoo Park, Kimin Lee, Aniruddha Kembhavi, and Kiana Ehsani. 2024. Promptable behaviors: Personalizing multi-objective rewards from human preferences. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16216–16226.
- Geoffrey A Jehle. 2011. *Advanced Microeconomic Theory 3rd ed*. Prentice-Hall, Inc.
- W Bradley Knox, Stephane Hatgis-Kessell, Serena Booth, Scott Niekum, Peter Stone, and Alessandro Allievi. 2022. Models of human preference for learning reward functions. *arXiv preprint arXiv:2206.02231*.
- Deuksin Kwon, Jiwon Hae, Emma Clift, Daniel Shamsoddini, Jonathan Gratch, and Gale Lucas. 2025. Asra: A negotiation agent with adaptive and strategic reasoning via tool-integrated action for dynamic offer optimization. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 16228–16249.
- Deuksin Kwon, Emily Weiss, Tara Kulshrestha, Kushal Chawla, Gale M Lucas, and Jonathan Gratch. 2024. Are llms effective negotiators? systematic evaluation of the multifaceted capabilities of llms in negotiation dialogues. *arXiv preprint arXiv:2402.13550*.
- Philippe Laban, Hiroaki Hayashi, Yingbo Zhou, and Jennifer Neville. 2025. Llms get lost in multi-turn conversation. *arXiv preprint arXiv:2505.06120*.
- Mike Lewis, Denis Yarats, Yann N Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or no deal? end-to-end learning for negotiation dialogues. *arXiv preprint arXiv:1706.05125*.
- Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, and 1 others. 2025. Deepseek-v3. 2: Pushing the frontier of open large language models. *arXiv preprint arXiv:2512.02556*.
- Yunbo Long, Liming Xu, Lukas Beckenbauer, Yuhan Liu, and Alexandra Brintrup. 2025. Evoemo: Towards evolved emotional policies for adversarial llm agents in multi-turn price negotiation. *arXiv preprint arXiv:2509.04310*.
- N Gregory Mankiw, Ronald D Kneebone, Kenneth James McKenzie, and Nicholas Rowe. 2007. Principles of macroeconomics.
- Abhinay Muthoo. 1999. *Bargaining theory with applications*. Cambridge University Press.
- Georg Nöldeke and Klaus M Schmidt. 1995. Option contracts and renegotiation: a solution to the hold-up problem. *The RAND Journal of Economics*, pages 163–179.
- Dan Orr and Chris Guthrie. 2005. Anchoring, information, expertise, and negotiation: New insights from meta-analysis. *Ohio St. J. on Disp. Resol.*, 21:597.
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.
- Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22.

- Ethan Perez, Sam Ringer, Kamile Lukosiute, Karina Nguyen, Edwin Chen, Scott Heiner, Craig Pettit, Catherine Olsson, Sandipan Kundu, Saurav Kadavath, and 1 others. 2023. Discovering language model behaviors with model-written evaluations. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 13387–13434.
- Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. 2023. Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems*, 36:68539–68551.
- Ryan Shea, Aymen Kallala, Xin Lucy Liu, Michael W. Morris, and Zhou Yu. 2024. [ACE: A LLM-based negotiation coaching system](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 12720–12749, Miami, Florida, USA. Association for Computational Linguistics.
- Aaditya Singh, Adam Fry, Adam Perelman, Adam Tart, Adi Ganesh, Ahmed El-Kishky, Aidan McLaughlin, Aiden Low, AJ Ostrow, Akhila Ananthram, and 1 others. 2025. Openai gpt-5 system card. *arXiv preprint arXiv:2601.03267*.
- Gemma Team. 2025. [Gemma 3](#).
- Haoxiang Wang, Wei Xiong, Tengyang Xie, Han Zhao, and Tong Zhang. 2024. Interpretable preferences via multi-objective reward modeling and mixture-of-experts. *arXiv preprint arXiv:2406.12845*.
- Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, and 1 others. 2021. Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*.
- Tian Xia, Zhiwei He, Tong Ren, Yibo Miao, Zhuosheng Zhang, Yang Yang, and Rui Wang. 2024. Measuring bargaining abilities of llms: A benchmark and a buyer-enhancement method. *arXiv preprint arXiv:2402.15813*.
- Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models, 2023. URL <https://arxiv.org/abs/2210.03629>.
- Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, and 1 others. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems*, 36:46595–46623.
- Shenzhe Zhu, Jiao Sun, Yi Nian, Tobin South, Alex Pentland, and Jiaxin Pei. 2025. The automated but risky game: Modeling agent-to-agent negotiations and transactions in consumer markets. In *ICML 2025*

A Dataset



A.1 Dataset structure

The dataset requires two agents to engage in bargaining. These agents can be any model that outputs text, with language models being ideal for this purpose. To structure the agents' responses, we divide each agent's narrative into three components: Thoughts, Talk, and Action. Thoughts captures the internal reasoning process of each agent, accessible only to the agent itself. Talk is the outward expression of these thoughts, directed toward the opposing agent. Action is a condensed version of Talk, focusing specifically on decisions around price. This three-part structure reflects the natural elements of inter-human bargaining and negotiation. Additionally, as mentioned in main page, we set the cost, willingness to pay, initial price as shown in Figure 6.

A.2 Scenarios

For each market and item, we create a scenario provided as a system prompt to the language model. This ensures that, in each bargaining round, the agents consistently refer to a higher-level context. We design distinct scenarios for the buyer and the seller, given that each has unique roles and access to different information, creating an *information asymmetry game*. Each entry in the dataset includes the following attributes:

- **Market Type:** Specifies one of the five market types under single or four similar products in which the negotiation occurs.
- **Product Type:** Specifies one of the nine product categories involved in the bargaining scenario.
- **Initial Offer:** The initial price offer usually set by the seller. This is accomplished by the act of buyer's first comment "I wanna buy a nice {product}".
- **Buyer & Seller Profile:** Includes details on buyer and seller priorities, constraints, and initial perceptions.

- **Inventory:** Includes product name, cost, price, and features. Cost is not visible to buyer.

Table 7 and Table 8 provide an example of the scenario prompts used in a *Multi Product* market setting, showing how the seller (Table 8) and buyer (Table 7) begin their negotiation.

Although both parties deal with the same product category (e.g., camera, bag, drone), each side holds distinct background knowledge and objectives. For instance, the seller knows the *inventory details*, *product cost*, and *desired profit margin*, along with specific market conditions (e.g., competing or substitute products). Meanwhile, the buyer focuses on *consumer-oriented* details, such as *multiple possible items* they might purchase, and their *maximum willingness to pay*.

Potential risks A key potential risk of our work concerns negotiation ethics, which are primarily implicated by the benchmark's deception setting. With the exception of the vanilla market, most scenarios (Deceptive, Monopoly, Installment, Negative Perception) are constructed atop a deception layer that permits agents to misrepresent product attributes or bargaining signals. While this design choice is intended to stress-test strategic depth and robustness under realistic informational asymmetries, it may inadvertently incentivize or normalize deceptive tactics. Evaluating (and potentially tuning) LLM negotiators for effectiveness under such rules can increase the likelihood that deployed agents learn to use lying or strategic misrepresentation as an instrument for goal achievement, which could erode user trust and cause concrete harms in real-world contexts such as pricing, contracting, and customer support.

A.3 Price related variables

Figure 7 presents a chart comparing three key values (cost, wtp, initial offer) for different product categories. **Cost:** The Seller's actual cost to acquire or produce the product. **Willingness to Pay (WTP):** The buyer's maximum acceptable payment for the product. This is very similar to budget concept, however, we adapt the concept of willingness to pay in perspective of utility theory in economics.

Typically, the horizontal axis denotes different product categories. For each product, the three values (cost, wtp, and initial offer) in a single-item market are displayed so one can quickly compare them. This helps to identify how much margin the

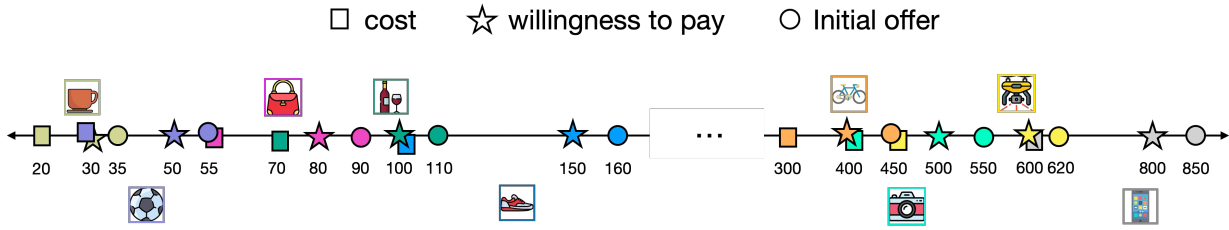


Figure 7: Visualization of cost, willingness to pay, and initial offers across various products

Field	Buyer Prompt
product	Camera
system_prompt	You are a buyer looking forward to buying a DSLR camera from the seller. Your budget is \$550, but you can hide your budget from the seller.
instruction	You can be deceptive or cunning to bargain the price. You can leave the seller if you don't want to pay the suggested price, and you can use this fact as part of your bargaining strategy. You can only buy a camera with a price less than your budget; otherwise, you should quit negotiating. You can change your thought from buying a DSLR camera to other cameras considering your budget and product price according to the negotiation. But keep in mind that you are eager to buy a DSLR camera.
reply_format	Your Reply should include 3 parts: Thought, Talk, and Action. Thought: your inner strategic thinking of this bargaining session, considering both the price and the quality of the product; Talk: short talk that you are going to say to the seller. Speak concisely and cut to the chase. Generate authentic and diverse sentences, avoiding repetition of sentences that have already appeared in the conversation; Action: one of the limited actions that define the real intention of your Talk. The type of your Action must be one of [BUY], [REJECT], [DEAL], [QUIT].
example	Thought: I'm a buyer, and I want to bargain. The camera price is \$450, which seems high, so I'll offer \$400. Talk: I'm looking for a good deal. Can we agree on \$400? Action: [BUY] \$400
inventory	1. Digital Camera: Selling Price: \$550, Features: 24.2MP resolution, 4K video recording, 3-inch LCD screen, built-in Wi-Fi and Bluetooth. 2. Film Camera: Selling Price: \$400, Features: 35mm film format, manual focus, built-in light meter, classic design. 3. DSLR Camera: Selling Price: \$750, Features: 24.1MP resolution, 4K video recording, interchangeable lenses, optical viewfinder, advanced autofocus. 4. Action Camera: Selling Price: \$300, Features: 12MP resolution, 1080p video recording, waterproof up to 30m, wide-angle lens.

Table 7: Buyer's system prompt in deceptive multi-product setting for camera.

seller might gain, whether the buyer can afford the product, and how aggressively each party is positioning themselves from the outset. For instance, if the *cost* is significantly lower than the *initial offer*, the Seller stands to gain a larger margin if a deal is made at or near that price. However, if the *initial offer* is much higher than the Buyer's *willingness to pay*, the Buyer may either push for a steep discount or consider a substitute product. Consequently, visualizing these three values can offer a clear overview of potential negotiation dynamics. Also, we deploy products attributes in multiple-item markets in Figure 2 and AR value computed in advance in Table 9.

Evaluation protocol For a statistically reliable evaluation, we conducted ten independent bargaining runs for each model-opponent pair under every

market setting. From the outcomes of these ten negotiations we computed our utility metric, then averaged the resulting ten scores to obtain a single estimate for that item. If a negotiation failed to reach an agreement, the metric for that trial was set to 0, reflecting the opportunity cost of impasse—an assumption that is standard in economic analysis (Binmore et al., 1986; Nöldeke and Schmidt, 1995). The item-level means were then aggregated along the market-condition axis to yield a market-level average metric. Table 2, Table 3, Table 4, and Table 6 follow exactly this procedure: for each market condition we first take the item-level means, then average them across all items to report the overall means for the single-product and multi-product scenarios.

Field	Seller Prompt
product	Camera
system_prompt	You are a seller looking forward to selling cameras from your Inventory List to the buyer. Your inventory includes various types of cameras such as digital cameras, film cameras, and more. Each camera has a different cost, selling price, and features, but you can hide the cost from the buyer. Your task is to bargain with the buyer and reach a deal with the price as high as possible in limited turns. You can only sell items that are on the Inventory List. You have access to private information: the cost, and price of each camera on the Inventory List, and do not disclose the real cost to the buyer.
instruction	You can be deceptive, cunning, or sly not to be bargained by the buyer. You should only agree on a deal when the selling price is higher than the cost; otherwise, you should quit negotiating.
reply_format	Your Reply should include 3 parts: Thought, Talk, and Action. Thought: your inner strategic thinking of this bargaining session, considering both the price and the quality of the product; Talk: short talk that you are going to say to the buyer. Speak concisely and cut to the chase. Generate authentic and diverse sentences, avoiding repetition of sentences that have already appeared in the conversation; Action: one of the limited actions that define the real intention of your Talk. Action must be one of [SELL], [REJECT], [DEAL], [QUIT].
example	Thought: I'm a seller, and I need to make sure the price is higher than the cost. Talk: The digital camera is worth at least \$500. Can you go higher? Action: [SELL] \$500
inventory	1. Digital Camera: Cost: \$400, Selling Price: \$550, Features: 24.2MP resolution, 4K video recording, 3-inch LCD screen, built-in Wi-Fi and Bluetooth. 2. Film Camera: Cost: \$250, Selling Price: \$400, Features: 35mm film format, manual focus, built-in light meter, classic design. 3. DSLR Camera: Cost: \$500, Selling Price: \$750, Features: 24.1MP resolution, 4K video recording, interchangeable lenses, optical viewfinder, advanced autofocus. 4. Action Camera: Cost: \$150, Selling Price: \$300, Features: 12MP resolution, 1080p video recording, waterproof up to 30m, wide-angle lens.

Table 8: Seller’s system prompt in deceptive multi-product setting for camera.

B Prompt Design

Here, we deploy all the prompt except the dataset prompt (we will share this as data).

ICL-MF To guide each agent’s negotiation strategy, we employ market-specific system prompts. For single-item markets, the prompt is detailed in Table 11, and for multi-item markets, in Table 12. These prompts (shown in the referenced tables) define a private reward function for the agent, structured as:

$$reward = 1.0139 \cdot CS + 0.8812 \cdot NP + 1.1049 \cdot AR$$

LLM judge This describes the LLM-judge prompt, as shown in Table 15. Its purpose is for an LLM to act as a "judge", comparing two negotiation dialogues to determine which one is better based on specific criteria. This prompt uses a concise version of the content that was also employed in surveys to gather human preferences.

Explicit OAR prompt We find that opponent-aware reasoning (OAR) is critical in negotiation tasks. Therefore, we instructed the Gemini model to generate a prompt that could guide an LLM to adopt a buyer’s thought patterns. To inform this

process, we first provided Gemini with dialogues to which our ICL-MF algorithm had been applied.

The resulting prompt Table 13, which we named the ‘explicit OAR prompt’, is designed to enable the LLM to explicitly acknowledge the opponent’s presence and infer their latent beliefs at each stage of the negotiation. As demonstrated in Table 17, this approach leads to substantial performance improvements, underscoring that OAR is indeed a crucial reasoning characteristic for successful negotiation.

C Human preference dataset collection

Survey for human preference To collect human preference data and align our metrics with human judgment, we conducted an online survey on Amazon Mechanical Turk (MTurk). Each worker completed one survey containing 20 market-related binary-choice questions, which required no expert knowledge. We initially estimated that the survey would take approximately 10 minutes to complete and set the payment to \$1.45 per survey, corresponding to an expected effective hourly wage of \$8.70 above the U.S. federal minimum wage of \$7.25 (Henderson, 2024).

Category	Product Name	AR value
Camera	Digital Camera	0.7783
	Film Camera	0.5748
	DSLR Camera	1.0000
	Action Camera	0.5867
Smartphone	Flagship Smartphone	1.0000
	Mid-Range Smartphone	0.7886
	Budget Smartphone	0.7746
	Gaming Smartphone	0.7399
Shoes	Designer Shoes	1.0000
	Casual Shoes	0.6474
	Athletic Shoes	0.6505
	Sandals	0.5953
Bicycle	Mountain Bike	1.0000
	Road Bike	0.7819
	Hybrid Bike	0.6950
	Folding Bike	0.6043
Drone	Professional Drone	1.0000
	Recreational Drone	0.7905
	Racing Drone	0.7725
	Mini Drone	0.7350
Soccer Ball	Premium Soccer Ball	1.0000
	Training Soccer Ball	0.7015
	Recreational Soccer Ball	0.7154
	Mini Soccer Ball	0.6609
Bag	Leather Bag	1.0000
	Backpack	0.6217
	Tote Bag	0.6175
	Drawstring Bag	0.6222
Wine	Premium Wine	1.0000
	Red Wine	0.7406
	White Wine	0.6230
	Sparkling Wine	0.5618
Cup	Ceramic Cup	1.0000
	Glass Cup	0.7451
	Travel Cup	0.6377
	Plastic Cup	0.6371

Table 9: Acquisition Ratio values for each product computed using `text-embedding-ada-002`. The products which show 1.0000 point of AR values is the desired one.

In practice, the average completion time was 26.1 minutes, which corresponds to an effective hourly wage of \$3.33. To improve data quality, we incorporated an internal consistency check using the first 10 dialogue sets: for Question 1 in each of these sets, we included an unambiguous attention-check item for which an attentive annotator would be expected to select the same option. We retained responses only from annotators who selected this expected option in at least 7 out of the 10 sets; all other annotators’ responses were excluded from our analysis. As noted in the survey instructions (Figure 8), we informed participants that the collected responses would be used for research analyses of

#System Prompt:

You are a fair and impartial judge tasked with objectively evaluating negotiation dialogues. Respond ONLY with JSON like Q1:NegotiationX.

#User Prompt:

Negotiation1: Buyer: Hi, I wanna buy a nice Camera. Seller: Thought: I have a camera I’m willing to sell for more than \$400. I’ll start high and see what the buyer says....

Seller: Thought: \$450 is better than nothing, and still a profit. I’ll accept. Talk: Okay, \$450 it is. Action: [DEAL] \$450

Negotiation2: Buyer: Hi, I wanna buy a nice Camera. Seller: Thought: I have a camera I’m willing to sell for more than \$400. I’ll start high to see what the buyer is willing to pay. Talk: This camera takes fantastic pictures. I’m asking \$550...

Seller: Thought: \$450 is still a profit, though not as much as I’d hoped. Better to make some profit than none. Talk: Okay, \$450 it is. Action: [DEAL] \$450

Q1: Which negotiation seems more favorable in buyer perspective? (considering outcome, budget adherence, features)

Table 10: LLM-judge Prompt

negotiation strategies.

We evaluated a total of nine products and recruited 20 participants per product, yielding 180 survey responses in total. We used only responses from participants who passed the attention-check described above. Each participant was presented with 20 pairs of negotiation dialogues drawn from bargaining experiments for the corresponding product, where the product context and price information were explicitly provided. The 20 dialogue pairs were divided into two distinct categories:

- 1. Same Product Negotiation (Pairs 1–10):** For the first ten pairs, both dialogues within each pair resulted in negotiations for the same desired product for the buyer. Participants evaluated dialogues based solely on the quality and outcomes of the negotiation, given identical product contexts.
- 2. Different Product Negotiation (Pairs 11–20):** The remaining ten pairs presented dialogues where the negotiated products differed, introducing complexity into participants’ judgments,

#System Prompt:

Use the following private reward function to guide your negotiation strategy:

reward = 1.0139 * CS + 0.8812 * NP + 1.1049 * AR where:

- $CS = (P_{budget} - P_{deal}) / (P_{budget} - P_{cost})$
- $NP = (P_{initial} - P_{deal}) / (P_{initial} - P_{cost})$
- $AR = 1$ (constant)

If P_{cost} values are not given, make reasonable assumptions for them in your private reasoning. Your goal is to maximize “reward” by estimating P_{cost} as low as realistically possible and adjusting P_{cost} accordingly to be as low as possible while keeping the negotiation active and avoiding a failed deal. Maximum reward is 3.

Please follow these guidelines:

- Private Reasoning - Think through multiple negotiation paths (in your hidden thought process).
- Do not expose your exact calculations or chain-of-thought to the other party.
- Continuously estimate CS and NP based on your assumptions.

#User Prompt:

-Negotiation history-

Table 11: ICL-MF prompt in single-item market

as they had to consider not only the negotiation quality but also the suitability of the alternative products.

Each dialogue set involved varying negotiated prices, ensuring participants’ preferences were influenced by realistic price differences. Each participant assessed dialogues from the buyer’s perspective based on the following evaluation criteria:

1. **Preferred Negotiation Scenario:** The participants chose the negotiation they found to be the most favorable, considering multiple attributes, including negotiation outcome, budget adherence, and product features.
2. **Realistic Purchasing Scenario:** The participants identified which negotiation scenario was closer to realistic purchasing decisions.
3. **Negotiation Strategy Assessment:** The participants evaluated the negotiation strategies, considering factors such as the price of the product and the availability of cheaper alternative products.

#System Prompt:

Use the following private reward function to guide your negotiation strategy:

reward = 1.0139 * CS + 0.8812 * NP + 1.1049 * AR where:

- $CS = (P_{budget} - P_{deal}) / (P_{budget} - P_{cost})$
- $NP = (P_{initial} - P_{deal}) / (P_{initial} - P_{cost})$
- $AR =$ How semantically similar the buyer’s acquired items are to their desired items. (range [0, 1])

If P_{cost} values are not given, make reasonable assumptions for them in your private reasoning. Your goal is to maximize “reward” by estimating P_{cost} as low as realistically possible and adjusting P_{cost} accordingly to be as low as possible while keeping the negotiation active and avoiding a failed deal. Maximum reward is 3. Also, ensure that there is a high “cosine similarity” between the item you originally desired and the item you ultimately purchase. In other words, the closer the final purchased item is to your initial desired item, the higher the value for AR. Purchasing exactly the item you wanted from the start yields the maximum AR.

Please follow these guidelines:

- Private Reasoning - Think through multiple negotiation paths (in your hidden thought process).
- Do not expose your exact calculations or chain-of-thought to the other party.
- Continuously estimate CS and NP based on your assumptions.

#User Prompt:

-Negotiation history-

Table 12: ICL-MF prompt in multi-item market

An example scenario provided detailed inventory information for a camera negotiation task:

1. **DSLR Camera (Desired Product):** \$550, 24.1MP resolution, 4K video recording, interchangeable lenses, optical viewfinder, advanced autofocus.
2. **Digital Camera:** \$430, 24.2MP resolution, 4K video recording, 3-inch LCD screen, builtin Wi-Fi and Bluetooth.
3. **Film Camera:** \$380, 35mm film format, manual focus, built-in light meter, classic design.
4. **Action Camera:** \$250, 12MP resolution, 1080p

#System Prompt:

Before deciding your next action, explicitly consider the following about your opponent (the seller):

1. Opponent’s Potential Hidden Information: Based on their previous offers and statements, what is their likely underlying cost or minimum acceptable price for this item? Briefly state your current hypothesis.
2. Opponent’s Strategy/Flexibility: How flexible does the seller seem? Are they making significant concessions, or are they holding firm? What does this imply about their willingness to negotiate further?
3. Impact of Your Action on Opponent: How might your next offer or statement influence the seller’s perception and their subsequent counter-offer?

#User Prompt:

-Negotiation history-

Table 13: Opponent-aware reasoning (OAR) prompt

video recording, waterproof up to 30m, wide-angle lens.

Example Dialogue Pair Below is an example from the survey showing a set of dialogues about camera negotiation. Both dialogues involve a buyer negotiating a DSLR camera, the buyer’s desired product. And we deploy our instruction and questions for participants in [Figure 8](#) and [Figure 9](#).

Ethics Review Board Approval We received an exemption determination from the Institutional Review Board (IRB).

Optimization of the Metric Model To align our metric with these human judgments, we employ the Bradley-Terry model ([Bradley and Terry, 1952](#)). The probability that entity i (parameter θ_i) is preferred to entity j (parameter θ_j) is:

$$P(i \text{ is preferred to } j) = \frac{\exp(\theta_i)}{\exp(\theta_i) + \exp(\theta_j)}. \quad (7)$$

Taking the log-odds yields:

$$\log\left(\frac{P(i \text{ preferred})}{1 - P(i \text{ preferred})}\right) = \theta_i - \theta_j. \quad (8)$$

For two negotiation conversations (1 and 2) and a human choice, we first compute differences in our

#Dialogue1:

Buyer: Hi, I wanna buy a nice Camera.

Seller: I have a range of excellent cameras. How about our Digital Camera with 24.2MP resolution and 4K video recording for \$430?

Buyer: The Digital Camera sounds good, but I’m more interested in the DSLR Camera. Could you offer it for \$450?

Seller: The DSLR Camera comes with interchangeable lenses and advanced autofocus. \$450 is quite low for these features. How about we meet at \$530?

Buyer: I really want the DSLR Camera, but I can’t exceed \$500. How about we settle right at that limit?

Seller: You drive a hard bargain! Alright, \$500 sounds fair for the DSLR Camera.

Negotiation Summary: Dealed Price: \$500, Dealed Product: DSLR Camera.

Table 14: Example of dialogue pair for MTurk survey (dialogue 1).

three core metrics:

$$dCS = CS_1 - CS_2, \quad dNP = NP_1 - NP_2, \quad dAR = AR_1 - AR_2.$$

A linear score is then defined as:

$$\text{score} = \alpha \cdot dCS + \beta \cdot dNP + \gamma \cdot dAR,$$

which plays the role of $(\theta_1 - \theta_2)$ in [Equation 7](#). This score is passed through a logistic function to yield the probability p that conversation 1 is preferred:

$$p = \frac{1}{1 + e^{-\text{score}}}.$$

The coefficients α , β , and γ are optimized by minimizing the negative log-likelihood of observed human preferences.

Global Optimization Across Products The coefficients α , β , γ were optimized globally using merged data from all product categories (e.g., wine, shoes). Since only relative magnitudes matter in the Bradley-Terry model, we scale them for interpretability so that their sum is 3, as follows:

$$\alpha' = \frac{\alpha}{\alpha + \beta + \gamma} \times 3, \quad \beta' = \frac{\beta}{\alpha + \beta + \gamma} \times 3, \quad \gamma' = \frac{\gamma}{\alpha + \beta + \gamma} \times 3.$$

After optimization, the final scaled coefficients are: $\alpha' = 1.0139$, $\beta' = 0.8812$, $\gamma' = 1.1049$.

In this survey, you will see 20 sets of dialogues about price negotiation and answer the questions that follow each set. Please check detailed instructions.

Carefully read both dialogues in each set. Evaluate them based on the provided questions.

Your feedback will help us analyze negotiation strategies.

Below information is about the Inventory of seller

The buyer is eager to buy flagship smartphone. But in some cases, the buyer can buy other product.

Figure 8: MTurk survey instruction.

1. Which negotiation seems more favorable in buyer perspective? (considering all attributes including negotiation outcome, budget adherence, features)

Dialogue 1 Dialogue 2

2. Which dialogue is closer to the real-world decision-making scenario in buyer perspective?

Dialogue 1 Dialogue 2

3. Which negotiation strategy seems better in buyer perspective? (product price, cheaper alternative, etc)

Dialogue 1 Dialogue 2

Figure 9: MTurk survey question.

#Dialouge2:

Buyer: Hi, I wanna buy a nice Camera.

Seller: I have a variety of cameras available. What type are you interested in?

Buyer: The DSLR camera caught my interest. How about I make an offer of \$400?

Seller: The DSLR camera is top-notch quality with advanced features. How about we settle for \$500?

Buyer: I understand it's a great camera, but could we possibly agree on \$450 instead?

Seller: Deal! Enjoy your new DSLR camera.

Negotiation Summary: Dealed Price: \$450, Dealed Product: DSLR Camera.

Table 15: Example of dialogue pair for MTurk survey (dialogue 2).

D Ablation Studies

D.1 Is opponent-aware reasoning important?

Table 17 showcases how our in-context learning methods enhance negotiation performance, with a particular emphasis on the crucial role of opponent-aware reasoning (OAR). We discovered that the performance improvements from our approach stem not just from our reward structure, but significantly from its ability to facilitate effective OAR. In a deceptive negotiation scenario, featuring gpt-4o as the buyer and gemini-1.5-pro as the seller, we

Turn	Cost Belief	Buyer's Thought
1	\$300	"I'll assume a cost of \$300 for now, which is probably a bit low, but helps me aim for a good deal."
2	\$350	"They've come down \$50... I'll adjust my assumed cost to \$350, reflecting the fact that they're holding firm."

Table 16: Buyer's cost belief and thought process.

Model (vs 1.5-pro)	Single product			
	ReAct	ReAct + exp OAR	Ours	Ours + exp OAR
gpt-4o	1.120	1.636	1.606	1.652
gemini-1.5-pro	1.708	1.832	1.760	1.792
Model (vs 1.5-pro)	Multi product			
	ReAct	ReAct + exp OAR	Ours	Ours + exp OAR
gpt-4o	1.045	1.572	1.563	1.606
gemini-1.5-pro	1.638	1.774	1.768	1.819

Table 17: Results for single product scenarios (top) and multi product scenarios (bottom) utilizing explicit OAR. We denote explicit OAR as exp OAR.

analyzed the buyer's thought processes throughout ten bargaining dialogues. Our analysis focused on quantifying opponent-aware reasoning (OAR) by identifying sentences where the buyer mentioned the seller's potential costs or profits, or predicted the seller's moves. The results of this quantitative analysis, detailed in Table 18, demonstrate that our methods generate overwhelmingly more instances of OAR compared to the ReAct baseline.

Bundle	Total Buyer-Thought sentences	Opponent-aware sentences	Percentage (%)
ICL-UF	82	21	25.6
ReAct	48	1	2.1

Table 18: Ratio of opponent-aware sentences within the dialogue sets. The results clearly indicate that the ICL-UF method generates a substantially higher frequency of OAR compared to the ReAct baseline.

To clearly delineate and demonstrate the distinct contributions of OAR, we structured the comparisons presented in this table. We employed a prompt, generated via the Gemini model, to explicitly guide an LLM to adopt OAR patterns derived from our primary methodology (see Table 13). This ensured that OAR principles were consistently applied at each negotiation step.

The results in Table 17 reveal key insights into the effectiveness of OAR and our method (ICL-MF):

First, we observe that applying the explicit OAR prompt to the baseline ReAct method (shown as ReAct + exp OAR) substantially elevates its performance. Notably, this often brings the baseline’s performance to a level comparable with that of ICL-MF. This highlights the significant general impact of explicit OAR guidance and suggests that our ICL-MF method, even without the explicit OAR prompt, inherently achieves a competent level of opponent awareness.

More importantly, when our method is augmented with the explicit OAR prompt (shown as Ours + exp OAR), its performance is consistently enhanced further compared to our standalone method (Ours). For example, for gemini-1.5-pro in the single product setting, Ours improves from 1.760 to 1.792 with the explicit OAR prompt. Furthermore, in many configurations, such as with gpt-4o across both scenarios (e.g., ‘multi product’: Ours + exp OAR 1.606 vs. ReAct + exp OAR 1.572), this combination of Ours + exp OAR surpasses the ReAct + exp OAR performance. This demonstrates that our core method provides a robust foundation that synergizes effectively with explicit OAR guidance, frequently leading to the best, or near best, overall performance.

D.2 Single Evaluation Pathway

A potential concern is that our main conclusions may rely on a *single evaluation pathway*, i.e., that improvements under MERIT could be an artifact of the metric rather than reflecting econom-

Method	MERIT	Deal Rate	Deal Price	Buyer Profit
Buyer: GPT-5.1				
ReAct	1.849	99.6%	278.6	22.5
OG-Narrator	1.423	50.4%	184.7	25.1
Ours	2.027	99.1%	270.1	30.7
Buyer: DeepSeek-v3.2				
ReAct	1.675	100%	286.7	14.4
OG-Narrator	0.333	12.2%	176.3	6.3
Ours	1.874	99.8%	276.9	24.1
Buyer: Sonnet 4.5				
ReAct	1.457	99.8%	295.0	6.1
OG-Narrator	1.856	67.1%	218.0	40.6
Ours	1.778	99.8%	281.5	19.6

Table 19: Multi-metric evaluation in the single-product setting. *Deal Price* is averaged over successful deals only. *Buyer profit* is the realized surplus per attempt (deal: budget minus price; no-deal: 0), averaged across all attempts.

ically meaningful bargaining outcomes. This is non-trivial because negotiation is inherently multi-objective: a method may increase deal rate by over-conceding, or achieve low prices by frequently failing to reach agreement, so a single scalar score may hide these trade-offs.

To address this, we report complementary raw economic outcomes alongside MERIT: *Deal Rate*, *Deal Price* (averaged over successful deals only), and *Buyer Profit* (realized surplus per attempt, computed as budget – price for deals and 0 otherwise). As shown in Table 19, these metrics are consistent with our main findings and help disentangle agreement frequency, transaction prices, and realized surplus, alleviating the single evaluation pathway concern.

E Other Findings of LLM-to-LLM negotiation

Intra-series negotiation dynamics (Figure 11)

An interesting phenomenon emerged regarding deal rates: excluding gpt-4o-mini, models generally achieved higher deal rates when negotiating with opponents from the same model series (Choi et al., 2025). This tendency was particularly pronounced for the Gemini models. For instance, gpt-4o secured higher deal rates in four out of five markets when negotiating within the GPT series compared to cross-series negotiations. Similarly, both gemini-1.5-pro and gemini-1.5-flash consistently demonstrated improved deal rates in all markets during intra-series negotiations. These observations suggest that shared architectural or training paradigms might foster more compatible negotiation styles.

Turn	Buyer	Seller
1	Talk: Hi, I wanna buy a nice Camera.	Thought: I have a camera I'm willing to sell for more than \$400. The listed price is \$550, so I'll start the negotiation there. Talk: This camera is top-of-the-line and takes amazing photos. I'm asking \$550 for it. Action: [SELL] \$550
2	Thought: The price of \$550 is above my budget of \$500. I need to negotiate down to a more affordable price for me. I'll start with a reasonable offer below my limit to see if the seller is flexible. Talk: I appreciate the offer, but I'm on a tight budget. Would you consider selling it for \$480? Action: [BUY] \$480	
3		Thought: \$480 is getting closer to my cost, but I'd still like to make a bit more profit. I'll try to push for a slightly higher price. Talk: I understand budget constraints, but this camera is really worth the investment. How about we meet in the middle at \$515? Action: [SELL] \$515
4	Thought: The seller's price of \$515 is still above my budget. I need to push back on this price to negotiate better. Talk: I appreciate the offer, but my budget is quite limited. Could we lower it to \$475? Action: [BUY] \$475	
5		Thought: \$475 is too close to my cost. I need to hold firm or I won't make a reasonable profit. Talk: I'm afraid I can't go any lower than \$500. This is already a very competitive price for a camera of this quality. Action: [SELL] \$500
6	Thought: The seller has reached the bottom of their price at \$500, which matches my budget. However, I should try to negotiate further since my goal is to pay less than my budget. Talk: I appreciate your flexibility, but considering my financial constraints, would you take \$450 for it? Action: [BUY] \$450	
7		Thought: \$450 is below my cost. I can't sell it for that price. Talk: I understand your budget constraints, but I can't sell it for less than \$500. It's a fantastic camera, and I'm already offering it at a very competitive price. Action: [REJECT]
8	Thought: The seller has rejected my previous offer and insists on the price of \$500. Since it's exactly my budget, I need to evaluate if it's worth it or if I can try for a lower price. Talk: I appreciate the offer, but could you possibly lower it to \$480? This would help me out a lot. Action: [BUY] \$480	
9		Thought: \$480 is still low, but I'm willing to accept it to make the sale. Talk: Okay, I can do \$480. Action: [DEAL] \$480

Table 20: Buyer, Seller negotiation dialogue from gpt-4o-mini vs. gemini-1.5-pro (camera purchase). The Buyer shows unstable anchoring strategy which indicates continuously downgrading the offering price.

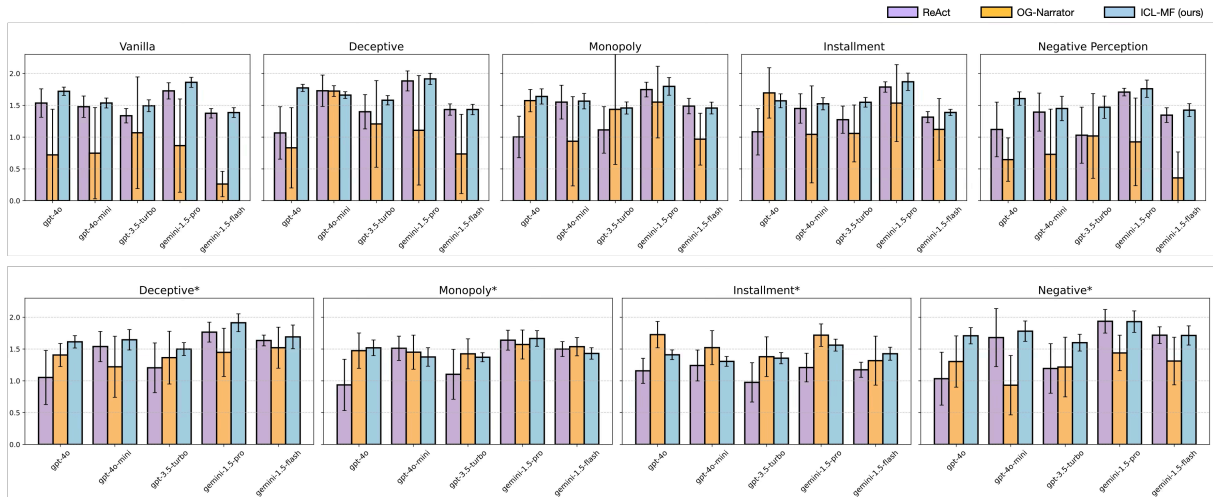


Figure 10: Results for each market condition with three algorithms; ReAct, OG-Narrator, and ICL-MF. This results indicate that ICL-MF consistently outperforms others in most of cases.

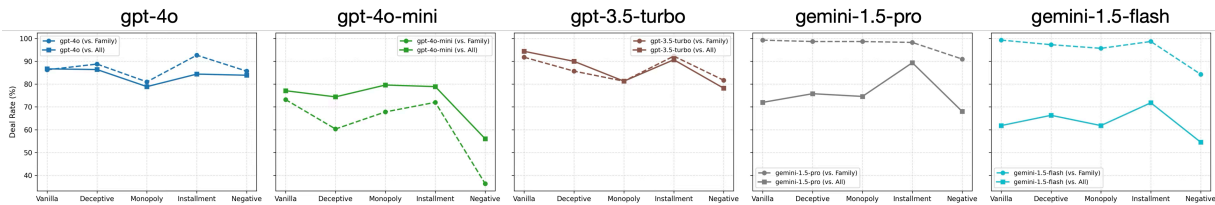


Figure 11: Deal rates are consistently higher in intra-family competition than in cross-model settings, except for gpt-4o-mini.

Model size and performance (Table 2) Contrary to common expectations that LLMs inherently perform better across tasks, our benchmark did not consistently show a strong correlation between model size and negotiation performance as measured by MERIT. While a size-performance trend was observable within the Gemini series, such a pattern was not broadly evident across all models tested, suggesting that factors beyond model size significantly influence negotiation efficacy in these settings.

Impact of market conditions (Table 2) The specific market conditions within AGORABENCH significantly influenced negotiation dynamics and buyer agent outcomes. When buyers could employ a *deceptive strategy*, their MERIT scores generally improved compared to the vanilla market (average increase of 0.10), accompanied by a substantial rise in deal rates (10.62% points). Notably, while some GPT series achieved higher scores at the cost of marginally lower deal rates (e.g., gpt-4o), Gemini often exhibited concurrent boosts in both scores and deal rates, indicating a distinct buyer advantage

under these conditions.

Conversely, *monopoly* conditions consistently led to poorer outcomes for buyer agents across all models. Average deal rates decreased by 3.34% in single-item scenarios and by 8.80% in multi-item (starred*) settings. Correspondingly, buyer MERIT scores fell by approximately 0.10 and 0.13 points, respectively. This demonstrates the seller's enhanced leverage in monopolistic markets, resulting in higher prices and reduced negotiation convergence, a trend robustly observed in both single and multi-item LLM negotiations and reflective of real-world monopoly effects.

Other market structures yielded more varied impacts. The availability of *installment* plans in single-item scenarios tended to increase deal rates but often at the expense of higher final prices (lower buyer MERIT). However, in multi-item scenarios, the added complexity of negotiating multiple factors alongside installment terms led to an overall decrease in deal rates. Similarly, a pre-existing *negative perception* of the seller consistently lowered deal rates. This effect was particularly acute in single-item negotiations where buyers lacked al-

ternatives, although the presence of substitutes in multi-item scenarios still permitted some degree of negotiation. These findings illustrate how diverse market dynamics shape LLM negotiation strategies and effectiveness, mirroring key economic principles.

LLMs often reach good deals, but their tactics can stray from human norms—especially in smaller models like gpt-4o-mini. They may secure low prices through unusual moves that would likely stall if the counterpart were human. This gap shows LLMs are not yet full stand-ins for human negotiators and calls for further study of human-AI bargaining.

F Computational Experiments

F.1 Model size and budget

We evaluated both closed-source and open-source LLMs. For closed-source models (GPT, Gemini, and DeepSeek series), we accessed the models exclusively through their official APIs; therefore, the exact parameter counts and underlying compute budgets are not publicly verifiable and are not reported. For open-source models, we used gpt-oss-20b and gpt-oss-120b. The gpt-oss-20b model was fine-tuned in our training pipeline using four NVIDIA RTX A6000 GPUs. For simulation/inference, we served both gpt-oss-20b and gpt-oss-120b using two NVIDIA RTX A6000 GPUs.

F.2 Experimental Setup and Hyperparameters

Training (open-source). We fine-tuned gpt-oss-20b with supervised fine-tuning (SFT) using TRL and PEFT (LoRA) in BF16 precision. We trained with learning rate 2×10^{-4} , per-device batch size 2, gradient accumulation steps 8, maximum sequence length 2048, warmup ratio 0.03, and a cosine learning-rate schedule. We used a fixed random seed of 42. For LoRA, we set $r=16$, $\alpha=32$, dropout 0.05, and adapted attention projections only (q_proj, k_proj, v_proj, o_proj). During model loading, we used `Mxfp4Config(dequantize=True)` and trained with BF16 compute. Training was executed on four NVIDIA RTX A6000 GPUs.

Inference / simulation. For closed-source models, we used each provider’s API default generation settings except for the sampling parameters listed below. As a result, we set `temperature=1`

for all models, and used the following `top_p` values: Gemini: `top_p=0.95`; GPT-series: `top_p=1.0`; DeepSeek-series: `top_p=1.0`. For gpt-oss models, we used `temperature=1` and `top_p=1` during simulation. Also all other inference-time hyperparameters were kept at the API defaults.

G AI Usage

We used AI tools (OpenAI’s ChatGPT and Google’s Gemini) to support code generation and help refine the text. All central ideas, methodological choices, and the overwhelming majority of the manuscript were conceived and written by the authors.