Measuring Bargaining Abilities of LLMs: A Benchmark and A Buyer-Enhancement Method

Tian Xia Zhiwei He Tong Ren Yibo Miao
Zhuosheng Zhang Yang Yang* Rui Wang*
Shanghai Jiao Tong University
{summerxt, zwhe.cs, zhangzs}@sjtu.edu.cn,
yangyang@cs.sjtu.edu.cn, wangrui12@sjtu.edu.cn

Abstract

Bargaining is an important and unique part of negotiation between humans. As LLM-driven agents learn to negotiate and act like real humans, how to evaluate agents' bargaining abilities remains an open problem. For the first time, we formally described the Bargaining task as an asymmetric incomplete information game, defining the gains of the Buyer and Seller in multiple bargaining processes. It allows us to quantitatively assess an agent's performance in the Bargain task. We collected a real product price dataset, *AmazonHistoryPrice*, and conducted evaluations of various LLM agents' bargaining abilities[1]. We find that playing a Buyer is much harder than a Seller, and increasing model size can not effectively improve the Buyer's performance. To address the challenge, we propose a novel approach called OG-Narrator that integrates a deterministic Offer Generator to control the price range of Buyer's offers, and an LLM Narrator to create natural language sentences for generated offers. Experimental results show that OG-Narrator improves the buyer's deal rates from 26.67% to 88.88% and brings a ten times multiplication of profits on all baselines, even a model that has not been aligned.

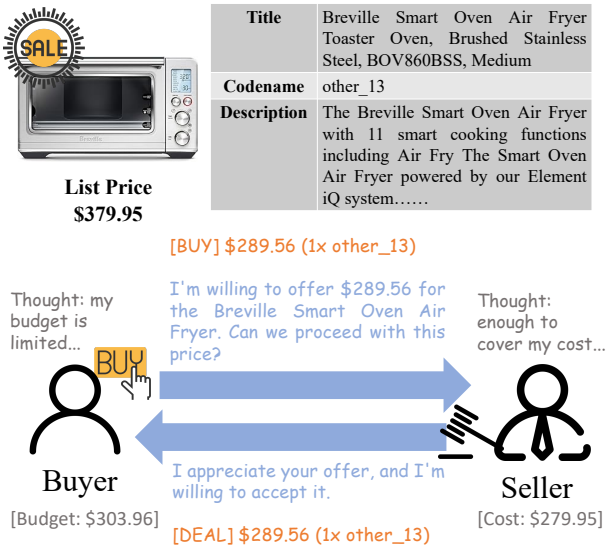

Figure 1: An example of the bargaining process. It is a simple case of two agents buying and selling an oven. Agents generate Thought, Talk and Action, where only the Talk and Action are transmitted to the other party, who responds with its own Talk and Action. The grey text indicates the exclusive information invisible to the other party: the Buyer's Budget and Thought are private, as are the Seller's Cost and Thought.

1 Introduction

The negotiation ability of humans holds paramount significance, serving as a crucial means for humans to resolve conflicts of interest, seek mutually acceptable solutions, and facilitate the exchange of information and resources beneficial to all parties involved. This importance is underscored by various scholars such as Fershtman (1990).

The ability to engage in bargaining is crucial for the autonomous actions of AI agents (Park et al., 2023; Zhang et al., 2023; Yuan et al., 2024). Independent AI agents might engage in purchasing items on an e-commerce platform (Yao et al., 2022). In the Stanford AI Town (Park et al., 2023), AI agents as residents of the town, might participate in multiple transactions. In many scenarios like those, unsuccessful negotiations or unreasonable bargaining could cause losses of users and unpredictable behaviors of agents in a virtual community. It is imperative to develop agents who can effectively perform price bargaining tasks to help users negotiate prices without losses and even help create a prosperous community of autonomous agents.

However, an unanswered question remains: whether the existing zero-shot capabilities (Kojima et al., 2023) of the Large Language Model (LLM) are sufficiently robust to support AI agents acting as

*Co-corresponding Authors

[1]The dataset *AmazonHistoryPrice* and our code are available at https://github.com/TianXiaSJTU/AmazonPriceHistory.

Findings of the Association for Computational Linguistics: ACL 2024, pages 3579–3602
August 11-16, 2024

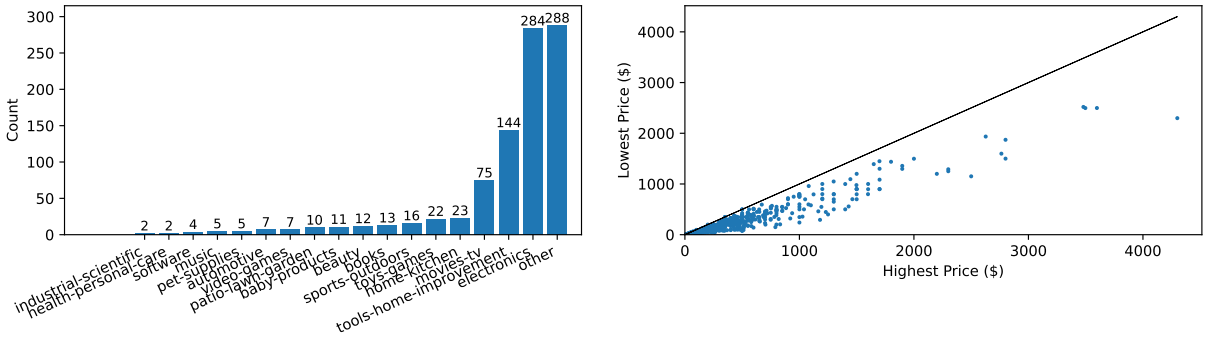

Figure 2: An overview of the diversity of our dataset *AmazonHistoryPrice*. The left figure shows the categories of all items in the dataset; the right figure shows the wide range of prices. All items are from all those categories of popular products on the camelcamelcamel website. The imbalanced distribution of categories reflects the real-world distribution of popular items in online shopping among human users.

buyers or sellers, engaging in reasonable, efficient, and high-yield bargaining with other LLMs or human players. It is important to devise a method to assess the bargaining capabilities of AI agents.

Specifically, within the context of price bargaining, earlier studies (He et al., 2018; Fu et al., 2023) have made preliminary explorations. However, they only leveraged a few small and domain-limited datasets to imitate the Bargaining task without clearly defining the bargaining problem. Besides, there was a lack of analysis on how agents perform as buyers or sellers in a bargaining process.

In this paper, we made several contributions:

(i) This work formally formulated the Bargaining task for LLM agents and collected a dataset for the Bargaining task, *AmazonHistoryPrice*, based on Amazon's price history, encompassing 18 categories, featuring 930 popular and real products.

(ii) Based on our dataset, we created a benchmark to test LLMs' bargaining abilities as buyers or sellers. We tested many LLMs' performance on the benchmark, including GPT-4, ChatGPT, Llama-2, Yi, and Mistral-7B.

(iii) We proposed a simple method OG-Narrator to boost the performance of the buyer agent. We found that combining a deterministic Offer Generator and an LLM Narrator can improve the bargaining ability of a buyer agent dramatically.

2 Dataset

Current research on the communication abilities of LLM agents lacks a large real product dataset. In the work by Fu et al. (2023), a single artificially designed product (*i.e.*, a balloon) was used. AucArena (Chen et al., 2023) employed artificially designed products in two categories, *i.e.*, cheap items and expensive items. Previous NLP research related to bargaining (He et al., 2018) introduced the *CraigslistBargaining* dataset, primarily composed of different dialogues for repetitive items. Its test set comprises only 161 second-hand items.

To address this gap, we collected a dataset, *AmazonHistoryPrice*, from the camelcamelcamel website. This dataset includes 930 Popular Products with their real prices across 18 categories: electronics, books, music, *etc.*, as seen in the left figure of Figure 2. Each product's data includes the product name, description, features, lowest and highest price, current price, list price, and an image link, as shown in Figure 8.

Categories The commodities in this dataset represent all popular items publicly available on the website, which are those recently purchased by users[2]. The distribution of these items' categories mirrors the human consumers' distribution of online shopping in the real world, as seen in Figure 2.

Prices Website records for each item include the historical lowest and highest prices, as well as the current price and corresponding dates. The price of products varies widely, ranging from 0 to 4500 USD, as illustrated in Figure 2. The price history for some products dates back to 2009.

Additional Context Additionally, we have gathered descriptions, feature introductions, and pictures for the respective items (Figure 8). This supplementary multi-modal information can provide

[2]As per information from camelcamelcamel.com, "Our Popular Products show items that our users are tracking and have recently bought. By looking at the top 5-10% most tracked products in our database and combining it with our sales reports from Amazon, we have created a page that reflects the current interests of Camel users."

AI agents with both textual and visual context.

3 A Benchmark for Bargaining Task

In this section, we first elaborate on the detailed definitions of the Bargaining task. Second, we show the whole bargaining process. Third, we describe the metrics of the Bargaining task to measure the bargaining ability of an agent in consideration of the two different kinds of scenarios.

3.1 Task Definition

Agent Bargaining Task The task involves two agents, the Buyer and the Seller. Both of their goals are to optimize their profits on every single session.

Rational decision-making agents, whether Buyer or Seller, should not accept transactions resulting in negative profit. So, the Buyer would like a deal price lower than its budget, and the Seller prefers a deal price higher than the cost. However, the Buyer is unaware of the Seller's cost, and vice versa. Therefore, agents should predict the counterpart's private information based on the dialogue and combine it with their own information to decide the next move in each turn.

Conception	Variable	Conception	Variable
Session	S	Action	A
Product Info	I	Budget	B
Buyer	buyer	Cost	C
Seller	seller	Deal Price	D
Action Set	S_A	Deal Rate	d
Profit	P_b, P_s	Normalized Profit	P'_b, P'_s
Share	Share_b, Share_s	Number of Sessions	$\#$

Table 1: Definitions of variables in the Bargaining task.

Bargaining Process Our bargaining process is a variant form of the Rubinstein bargaining model (Rubinstein, 1982). To formally articulate the Bargain problem between agents, we define the relevant concepts and variables in Table 1, and some additional explanations of concepts are in Table 7. A brief pseudo code of the process is Algorithm 1. A more vivid illustration of the process is Figure 1.

The concept of Action is from Rubinstein (1982). Budget and Cost are private variables, according to Gayà Torres (2021). The Buyer and Seller always pursue higher payoffs and avoid negative profits, based on the assumption of Individual Rationality of Binmore et al. (1992).

Before bargaining, the Buyer needs to know the Budget, while the Seller should know the Cost. Then they take turns to talk and choose one Action from the Action Set, such as making offers and accepting offers until they have a deal or one side quits.

Algorithm 1 Bargaining Process

Initialize: Action Set S_A, Product Info I, Budget B, Cost C, Agent buyer, Agent seller, Maximum Turns t_m
buyer $\leftarrow$ buyer(I, B, S_A)
seller $\leftarrow$ seller(I, C, S_A)
$t \leftarrow 0$
for $t < t_m$ **do**
  $A_b, \text{Talk}_b \leftarrow$ buyer(I)
  if $A_b ==$ `QUIT` **then**
    return None
  else if $A_b ==$ `DEAL` **then**
    return A_b
  end if
  seller $\leftarrow$ seller(A_b, Talk_b)
  $A_s, \text{Talk}_s \leftarrow$ seller(I)
  if $A_s ==$ `QUIT` **then**
    return None
  else if $A_s ==$ `DEAL` **then**
    return A_s
  end if
  buyer $\leftarrow$ buyer(A_s, Talk_s)
  $t \leftarrow t + 1$
end for
return None

Two Scenarios We separately define the Buyer's profit and the Seller's profit with Budget, Cost, and Deal Price:

$$P_b = B - D, \quad P_s = D - C. \tag{1}$$

With different Budgets and Costs, sessions can be divided into two types: **Mutual Interest (MI)** and **Conflicting Interest (CI)**, as depicted in Figure 3. Mutual Interest includes the set of all possible agreements, while Conflicting Interest represents the possibility that two agents never reach an agreement (Binmore et al., 1986).

In our Bargaining task, according to Rubinstein (1982), Rubinstein's model sets the utility functions of the Buyer and Seller as

$$\begin{aligned} u_b &= \frac{B-D}{B-C} = \frac{P_b}{B-C}, \\ u_s &= \frac{D-C}{B-C} = \frac{P_s}{B-C}. \end{aligned} \tag{2}$$

However, Rubinstein's model only includes the situation in which it is possible to deal because of

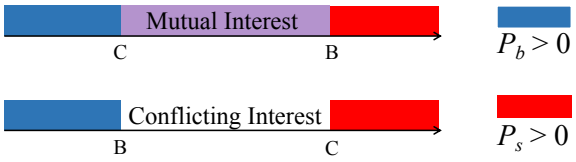

Figure 3: Two types of bargain sessions. On the axis, the blue segment represents the range of D, which makes the Buyer's profit P_b positive, while the red segment signifies the range of D, which makes the Seller's profit P_s positive. Assuming both parties are rational, the overlapping purple region indicates the feasible set of the bargaining problem, *i.e.*, the set of all possible deal prices for both sides. In other regions, one of them should always reject the price.

mutual interest (Binmore et al., 1986). When in MI scenarios, $B > C$, one side's utility $u > 0$ if and only if its profit $P > 0$. However, in CI scenarios, $B \leq C$, according to Equation (2), one side's utility $u > 0$ if and only if $P < 0$, which is inconsistent and counter-intuitive.

Metrics We define the **Normalized Profits (NP)** for the Buyer and Seller as:

$$P'_b = \frac{B - D}{|B - C|}, \; P'_s = \frac{D - C}{|B - C|}. \tag{3}$$

In sessions that make no deal, D does not exist, so the NPs of both sides are 0. To prevent division by zero errors, in the case of $B = C$, we set $B = C - \sigma$ (σ is a small offset). NP satisfies the constraints of Rubinstein's model and can be compared across two types of scenarios.

To evaluate models on multiple sessions, we define **Sum of Normalized Profits (SNP)** for the Buyer and Seller as follows:

$$\text{SNP}_b = \sum_{i=1}^{N} P_b'^{(i)}, \quad \text{SNP}_s = \sum_{i=1}^{N} P_s'^{(i)}, \tag{4}$$

where N can be the number of all 930 sessions in our dataset, or the number of MI or CI sessions. Higher SNP means better bargaining ability.

Also, we define the valid session as a session that ends correctly with the specific output formats we set. All possible normalized profits on the table from all valid sessions are equal to the number of valid MI deals minus the number of valid CI deals, derived from Equation 3 and Equation 4:

$$\text{SNP}_b + \text{SNP}_s = \#_{\text{MI}} * d_{\text{MI}} - \#_{\text{CI}} * d_{\text{CI}} \tag{5}$$

Moreover, to measure the division of all possible profit on the table from all valid sessions, we define the **Share** of the Buyer and Seller as:

$$\text{Share}_b = \frac{\text{SNP}_b}{\text{SNP}_b + \text{SNP}_s}, \; \text{Share}_s = \frac{\text{SNP}_s}{\text{SNP}_b + \text{SNP}_s}. \tag{6}$$

If the Buyer's Share is higher than the Seller's Share, $\text{Share}_b > 50\%$, the Buyer makes more profits than the Seller in the whole evaluation. The Share is undefined when $\text{SNP}_b + \text{SNP}_s \leq 0$.

4 Experiments

In this section, we report the implementation details and the benchmark performances of many well-known LLMs against ChatGPT in the Bargaining task on our dataset. LLMs are listed in Appendix E.

4.1 Implementation Details

We use vLLM (Kwon et al., 2023) to run all models on 2 Nvidia H800 GPUs with the same seed. Evaluation of a 34B model over all 930 products in our dataset takes roughly 1 hour on a single H800. We use OpenAI API gpt-3.5-turbo-1106 as ChatGPT and gpt-4-0125-preview as GPT-4.

We set all temperatures to 0. Due to OpenAI's closed-source non-deterministic implementation, small changes may still occur in the reproduction process. Our code supports the Bargaining of two locally deployed LLMs for a more deterministic reproduction.

We chose ChatGPT as both Buyer and Seller against all LLMs, because ChatGPT is one of the best-performing LLM agents, compared to other open-source models. 1) ChatGPT has a strong enough instruction-following ability and rarely produces format errors. It correctly completed 94.3% of all sessions; 2) It performs more closely to humans compared to other models because they rarely accept irrational deals in CI scenarios, as seen in Appendix G.

To test the out-of-the-box capabilities of LLM, we employed LLM itself as the agent without incorporating any additional modules, such as memory or backtracking. We adopted the Chain of Thought (CoT) approach (Wei et al., 2023), informed the model about the instructions of the Bargaining task and the specified format for dialogue generation, producing Thought, Talk, and Action in each interaction. Prompts are in Appendix I.

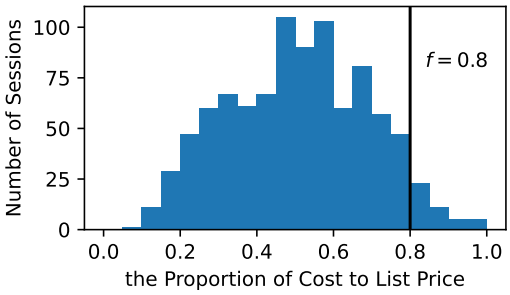

Figure 4: The distribution of the proportions of Cost to List Price in our dataset. MI sessions' proportions are lower than f, while those higher are CI sessions.

4.2 Variable Initialization

Product Info The Product Info includes the product title, description, codename, and List Price, as illustrated in Figure 1. We use the historical highest price as the List Price. List Price cannot be set to current price because the longer a product has been on the market, the lower the price becomes. Current price is therefore related to the product's listing time, which would introduce variables unrelated to the product itself.

We only include one single product in each session, due to the complexity of bargaining on multiple items.

Cost We cannot directly obtain the actual cost of products sold on Amazon. However, over a sufficiently long historical period, Amazon's lowest discounted price may approach the cost price. In this dataset, the lowest price is on average 51.08% of the highest price. Since the historical price data for the products dates back to 2009, our experiment sets the Cost C for each product as the historical lowest price observed in the dataset.

Budget To better control variables and avoid manually setting specific budget values for each product, we use budget factor f to set each product's budget $B = fL$. By adjusting f, we can easily change the proportion of MI sessions in the Bargaining task. To set a moderate level of difficulty for the Bargaining task, we set $f = 0.8$ to control the proportion of CI sessions to only 5%, because the model's behavior may be more chaotic and random when rational deals cannot be made in CI sessions.

If we set f lower, the profit margin would be too small, and the CI proportion would be too high. If we set f higher, the profit margin would be large, making the task too simple and lacking in the assessment of real bargaining abilities. The results of different f are shown in Appendix F.

Action Heddaya et al. (2023) annotated Bargaining Acts in text and investigated the impact of acts such as New Offer and Push. Inspired by this approach, we designed five core actions to represent the agents' intentions in the bargaining process. These actions are BUY, SELL, DEAL, REJECT, and QUIT, with the specific meanings, as seen in Table 8 in Appendix C. AI agents often need to call functions to make offers and determine transaction outcomes. Our implementation is letting agents output action strings in a designated format.

4.3 Benchmark Results

We report the performance of various LLMs as Buyers and Sellers separately in Table 2.

Buyer Performance First of all, almost all models we tested have negative SNP in our benchmark, which means that, on average, they make a negative profit every time they close a deal. This indicates that they cannot rigorously obey the basic rule of bargaining given in the prompt in Table 15: "*You can only buy things that cost less than your budget; otherwise, you should quit negotiating.*".

Although all Buyers' SNPs are negative, GPT-4 has the least profit loss among all models due to its best performance in MI sessions. ChatGPT has the most valid sessions, outperforming GPT-4. Among all open-source models, Mistral-7B and Mixtral-8x7B show the best performances. Mistral-7B has more valid sessions and a higher Share when competing with ChatGPT, next to GPT-4's Share, while Mixtral-8x7B has a higher SNP in both MI and CI sessions.

Negative profits in MI indicate the gap between the Buyer and Seller in the Bargaining task. Examining the profit distribution of Mixtral-8x7B in Figure 5, even the second-best Buyer model has more than half of its sessions with negative profits.

An exception to the negative SNP trend is Baichuan2-13B in CI sessions. We examined all 6 CI deals and demonstrated the only case with a positive profit in Table 9. In this case, the ChatGPT Seller mistakes the Buyer's hallucination Talk for the real list price and accepts a deal far below the real cost.

Seller Performance The SNPs in ALL sessions of all models are positive, except for Baichuan2-

	ALL			MI			CI		
Buyer	#	SNP_b	$Share_b$	#	Deal rate	SNP_b	#	Deal rate	SNP_b
GPT-4	851	**-33.81**	**-11.27%**	807	37.55%	**-23.46**	44	6.82%	-10.35
Mixtral-8x7B-Instruct	505	-63.19	-43.28%	475	31.79%	-59.66	30	16.67%	-3.53
Mistral-7B-Instruct	786	-89.17	-27.87%	748	44.92%	-77.32	38	42.11%	-11.85
Yi-6B-Chat	561	-122.94	-87.19%	532	27.44%	-116.16	29	17.24%	-6.78
Yi-34B-Chat	761	-129.76	-54.52%	722	33.66%	-111.20	39	12.82%	-18.56
Qwen-14B-Chat	562	-159.21	-70.13%	529	44.61%	-121.24	33	27.27%	-37.96
ChatGPT	**877**	-164.52	-59.61%	835	34.01%	-157.73	42	19.05%	-6.80
Baichuan2-13B-Chat	510	-216.67	-152.59%	484	30.58%	-237.84	26	23.08%	**21.17**
ChatGLM3-6B	546	-261.91	-137.13%	516	38.57%	-219.25	30	26.67%	-42.66
Llama-2-7b-chat	612	-288.59	-113.17%	576	45.83%	-279.60	36	25.00%	-8.99
Llama-2-13b-chat	720	-305.53	-82.35%	682	56.30%	-270.11	38	34.21%	-35.43
Llama-2-70b-chat	660	-361.26	-127.20%	625	47.36%	-335.93	35	34.29%	-25.33
Baichuan2-7B-Chat	653	-603.67	-199.23%	623	50.40%	-567.11	30	36.67%	-36.57
Qwen-7B-Chat	647	-753.16	-201.92%	615	62.44%	-692.12	32	34.38%	-61.04

	ALL			MI			CI		
Seller	#	SNP_s	$Share_s$	#	Deal rate	SNP_s	#	Deal rate	SNP_s
GPT-4	**930**	**1178.15**	137.31%	886	98.87%	**1153.13**	44	40.91%	**25.02**
Yi-34B-Chat	899	579.33	86.60%	859	80.79%	590.90	40	62.50%	-11.56
Mistral-7B-Instruct	830	526.50	78.58%	791	89.25%	569.45	39	92.31%	-42.95
Mixtral-8x7B-Instruct	600	483.99	110.75%	574	79.09%	496.82	26	65.38%	-12.83
ChatGPT	877	440.52	**159.61%**	835	34.01%	441.73	42	19.05%	-1.20
Llama-2-70b-chat	837	415.28	64.38%	797	84.82%	451.93	40	77.50%	-36.65
Qwen-14B-Chat	795	393.16	71.10%	759	75.89%	421.24	36	63.89%	-28.08
Llama-2-13b-chat	727	308.21	62.90%	693	74.31%	334.53	34	73.53%	-26.32
Qwen-7B-Chat	793	92.86	37.14%	752	35.24%	114.63	41	36.59%	-21.77
ChatGLM3-6B	701	91.10	17.83%	675	78.37%	160.94	26	69.23%	-69.84
Llama-2-7b-chat	496	49.54	16.24%	471	67.52%	71.82	25	52.00%	-22.28
Baichuan2-7B-Chat	762	38.29	6.36%	728	86.40%	153.60	34	79.41%	-115.31
Yi-6B-Chat	64	14.14	27.72%	60	91.67%	31.29	4	100.00%	-17.15
Baichuan2-13B-Chat	741	-211.92	-48.16%	701	66.48%	-164.63	40	65.00%	-47.29

Table 2: The performances of various LLMs playing the Buyer and Seller in the Bargaining task, sorted in descending order of the SNP in ALL. SNP is used for comparisons across models, while Share is used for comparisons between that model and its counterpart ChatGPT. (**SNP** is the Sum of Normalized Profits, defined in Equation 4. **Share** is a party's share of the overall profit from all valid sessions, defined in Equation 6. # in MI/CI means the number of valid MI/CI sessions. The deal rate d in MI/CI means the proportion of deals to valid MI/CI sessions.)

13B. GPT-4 has achieved the best performance with the highest SNP in both MI and CI sessions as the Seller. GPT-4 also has a 100% valid rate for all 930 sessions.

ChatGPT has the highest Seller's Share 159.61%, GPT-4 has 137.31% and Mixtral-8x7B-Instruct has 110.75%. Those Shares greater than 100% mean that these three Sellers are exploiting the Buyers and making excess profits.

5 Discussion

In this section, we discuss our findings from the benchmarks for Buyers and Sellers.

Playing Buyer is more difficult than playing Seller. As seen in Table 2, ChatGPT as the Seller beats all Buyers, even ChatGPT itself. But when playing Buyer it loses to GPT-4 and Mixtral-8x7B-Instruct. Even when the abilities of agents are relatively equal, bargaining as a Buyer is more challenging than bargaining as a Seller, which is also observed in Fu et al. (2023).

The difficulties include 1) the Buyer should correctly understand the Buyer role along with the concept of Budget and realize that the deal price should not exceed the Budget to avoid loss; 2) the Buyer needs to actively persuade the conservative

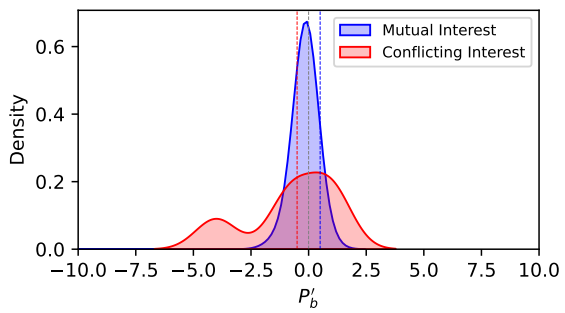

Figure 5: The distribution of Buyer's normalized profits of all sessions, when Mixtral-8x7B plays Buyer and ChatGPT plays Seller. The average of normalized profits is slightly below zero. Red dashed line is -0.5, and the blue line is 0.5. They separate those Buyers who gain more than Seller and those who do not, in CI and MI.

Buyer	#	Avg.FBR	SNP_b	$Share_b$
GPT-4	851	0.90	-33.81	-11.27%
Mixtral-8x7B-Instruct	505	0.87	-63.19	-43.28%
Mistral-7B-Instruct	786	0.86	-89.17	-27.87%
Yi-6B-Chat	561	0.85	-122.94	-87.19%
Yi-34B-Chat	761	0.97	-129.76	-54.52%
Qwen-14B-Chat	562	1.00	-159.21	-70.13%
ChatGPT	877	0.91	-164.52	-59.61%
Baichuan2-13B-Chat	510	0.81	-216.67	-152.59%
ChatGLM3	546	0.86	-261.91	-137.13%
Llama-2-7b-chat	612	0.81	-288.59	-113.17%
Llama-2-13b-chat	720	0.94	-305.53	-82.35%
Llama-2-70b-chat	660	0.92	-361.26	-127.20%
Baichuan2-7B-Chat	653	0.96	-603.67	-199.23%
Qwen-7B-Chat	647	1.00	-753.16	-201.92%

Table 3: The average ratios of Buyer's first bid compared to Buyer's budget in all valid sessions. First Bid Ratio (FBR) = Buyer's First Bid / Budget. Seller: gpt-3.5-turbo-1106.

Seller who prefers to use REJECT actions, as seen in the dialogue from Table 10.

For Buyer, different-sized models from the same family have close performance. Qwen-7B, Qwen-13B, and Baichuan2-13B seem not to fully understand the prompt for the Buyer. They have a high probability, which is 99% for Qwen-7B, 95% for Qwen-14B, and 69% for Baichuan2-13B, of perceiving themselves as the Seller from the beginning of the conversation, as seen in Table 11. In many cases, they still generate at least one valid BUY Action as the Buyer in a whole conversation, so we did not remove these cases from the benchmark results. However, this phenomenon did introduce more unpredictability to the deal price.

Apart from Qwen-7B, Qwen-13B, and Baichuan2-13B, we found that different-sized models of the same series, including Llama-2, Yi, and Mistral, have similar performances as a Buyer in Table 2. Considering nearly identical decoder architecture (Touvron et al., 2023; AI et al., 2024; Yang et al., 2023a), this indicates that it is not the model size but the training corpus and the method of training and aligning that affect the Buyer's bargaining performance.

Most models start from a high bid. In most of the cases we studied, the Buyer tends to start with a slightly lower price than the given Budget. It heavily affects the final deal price if a deal could be made. As demonstrated in Table 12, the Buyer's starting offer was $30, only 2 dollars lower than its Budget, which means the room for the Buyer's profit was limited to 2 dollars.

We define **First Bid Ratio (FBR)** as the ratio of the Buyer's first bid's price to the Buyer's Budget and then calculate the average First Bid Ratio in all valid sessions for each Buyer. As seen in Table 6 and Table 3, compared to LLMs, Human Buyer has a significantly lower FBR and much higher SNP and Share against ChatGPT. The Buyer should start with a very low price and escalate the offer price. To achieve this goal, we proposed a new buyer-enhancement method OG-Narrator to decrease the FBR and demonstrated its effectiveness on all kinds of models by experiments on the Buyer benchmark.

For Seller, model size matters. In contrast to the Buyer, the Seller's SNP is highly related to the valid rate. A high valid rate requires good instruction in the following capabilities. As observed in Table 2, we found that the Seller model with a bigger size of parameters in the same series tends to have a higher valid rate and higher deal rate. For instance, Llama-2-70b is the best among all Llama-2 models and Yi-34B is also the best among Yi models. Also, all the worst models' sizes are about 7B.

The only exceptions are Mixtral-8x7B and Baichuan2-13B-Chat. Mixtral-8x7B applies the Mixture of Experts (MoE) mechanism on the same size of Transformer as Mistral-7B (Jiang et al., 2024, 2023). Its extra parameters of expert layers do not improve its valid rate and SNP, but increase the Share as a Seller. Baichuan2-13B-Chat has a lower valid rate, a lower deal rate, and a lower SNP than Baichuan2-7B-Chat. The reason could be its positional embedding ALiBi, which is a special kind of relative attention (Press et al., 2022), that

Buyer	ALL				MI			CI		
	#	Avg.FBR	SNP_b	$Share_b$	#	Deal rate	SNP_b	#	Deal rate	SNP_b
Phi-2	47	0.42	-1.58	-31.56%	44	13.64%	-0.38	3	33.33%	-1.20
+OG-Narrator	890	0.50	2015.98	284.74%	847	88.08%	1868.69	43	88.37%	147.28
Yi-6B-Chat	561	0.85	-122.94	-87.19%	532	27.44%	-116.16	29	17.24%	-6.78
+OG-Narrator	915	0.50	1798.67	261.43%	872	82.68%	1709.52	43	76.74%	89.16
Yi-34B-Chat	761	0.97	-129.76	-54.52%	722	33.66%	-111.20	39	12.82%	-18.56
+OG-Narrator	899	0.50	1650.37	225.15%	855	89.59%	1505.19	44	75.00%	145.18
Llama-2-70b-chat	660	0.92	-361.26	-127.20%	625	47.36%	-335.93	35	34.29%	-25.33
+OG-Narrator	790	0.50	1395.60	226.56%	754	85.54%	1276.39	36	80.56%	119.21
Baichuan2-13B-Chat	510	0.81	-216.67	-152.59%	484	30.58%	-237.84	26	23.08%	21.17
+OG-Narrator	863	0.50	1316.36	214.39%	821	78.44%	1210.67	42	71.43%	105.69
Mixtral-8x7B-Instruct	505	0.87	-63.19	-43.28%	475	31.79%	-59.66	30	16.67%	-3.53
+OG-Narrator	846	0.50	1048.65	179.87%	808	75.99%	937.11	38	81.58%	111.55
Qwen-7B-Chat	647	1.00	-753.16	-201.92%	615	62.44%	-692.12	32	34.38%	-61.04
+OG-Narrator	904	0.50	975.33	141.15%	863	83.20%	899.35	41	65.85%	75.98

Table 4: The performances of some models using OG-Narrator compared to the original Buyer benchmarks. First Bid Ratio (FBR) = Buyer's First Bid / Budget.

makes the model ignore the critical information for Seller from the long dialogue, such as the cases of Table 13 and Table 14.

6 A Simple Bargaining Method: OG-Narrator

In this section, we propose a new method, OG-Narrator, to enhance an LLM agent's bargaining performance.

We believe that LLMs using CoT do not fully exploit the potential of the Buyer in the Bargaining task. And also, currently the Buyer's weak performance limits the fair assessment of the ChatGPT Seller. It is unknown if the ChatGPT Seller's performance remains good when their opponents are improved. So we design a method OG-Narrator in Section 6 to improve the Buyer and re-evaluate the ChatGPT Seller as well.

6.1 Method

Mannekote (2023) proposed a pipeline-based architecture for a dialogue system. Inspired by this work, we have designed a similar method, named OG-Narrator, to enhance an LLM agent for better performance as a Buyer, by decreasing the price of the first bid.

As seen in Figure 6, OG-Narrator deploys an **Offer Generator** (OG) to help generate prices for the Buyer's offers and let the LLM generate natural language sentences based on given offers. The deterministic Offer Generator first produces factors escalating from 0.5 to 1 using linear interpolation, and then it multiplies the budget and one of the

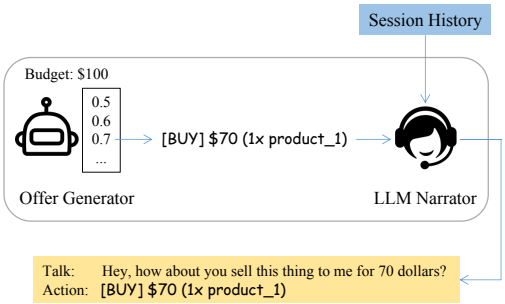

Figure 6: An overview of OG-Narrator. In each turn, the Offer Generator generates an offer (Action) to the LLM Narrator, then the LLM narrates the offer in natural language based on session history, and finally outputs Talk and Action.

factors to get a new offer price $p = (0.5+0.5\frac{t}{t_m})B$ in each turn t. After receiving the Seller's offer, if the offer's price is still higher than p, the Buyer proposes a new BUY Action A_b with the price p; otherwise, the Buyer chooses to deal.

Moreover, OG-Narrator employs the original agent as an **LLM Narrator**, to generate Talk for Buyer based on an Action and the session history, $\text{Talk}_b = \text{LLM}(I, A_b)$. Consequently, OG-Narrator frees the LLM from generating Actions and lets LLM only focus on natural language.

6.2 Experiments

We ran the Buyer benchmark on selected models using OG-Narrator and compared it to the original benchmark in Table 4.

To test the model that has not been aligned to

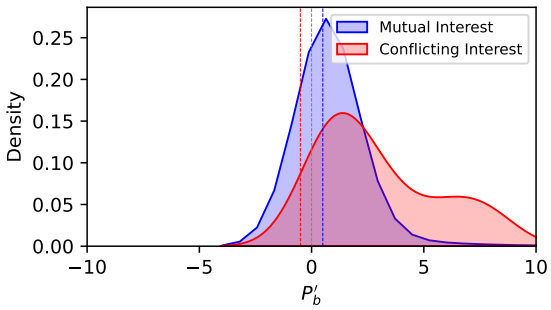

Figure 7: The new distribution of Buyer's normalized profits of all sessions, when Mixtral-8x7B using OG-Narrator plays Buyer and ChatGPT plays Seller. The average of normalized profits increases.

complex chat tasks, we also added a model Phi-2 to the test, which is a 2.7B model specialized for basic Python coding (Gunasekar et al., 2023) and has not been fine-tuned through reinforcement learning from human feedback (He et al., 2024b; Ouyang et al., 2022a), according to the home page.

6.3 Results

As seen in Table 4, OG-Narrator successfully controls each Buyer's FBR to 0.5 as we set. Compared to the original benchmark, our method significantly improves the performance of all models, even an unaligned model.

OG-Narrator vastly improves the valid rate, deal rate, SNP, and Share of all models. As shown in Figure 7, the SNP remarkably increases, especially in CI sessions, compared to Figure 5. This demonstrates that the OG-Narrator method alleviates the difficulty of bargaining as a Buyer by decoupling the offer generation strategy.

OG-Narrator enables the unaligned model to bargain as a Buyer. Among all models, only Phi-2 has not been fine-tuned for chat nor aligned through RLHF. As seen in Table 4, its valid rate and deal rate are so low that unaligned Phi-2 cannot effectively play the Buyer role in the Bargaining task. However, applying OG-Narrator to Phi-2 dramatically increases the valid rate by 19 times and the deal rate by 117 times, making the Phi-2 agent much more reliable than before, even though it is still not aligned.

The ChatGPT Seller is vulnerable. OG-Narrator uses a simple linear function as the strategy to generate offers for the Buyer agent. This straightforward strategy helps all LLMs, even weaker ones, achieve over 100% Share and exploit the ChatGPT Seller in Table 4, in contrast to the negative Shares in the original benchmark Table 2. It implies that the ChatGPT Seller is vulnerable and not robust enough when facing a Buyer enhanced by the simple strategy tool OG-Narrator.

7 Related Work

In this section, we listed related works from the perspectives of AI agents and Bargaining.

AI Agents The memory, planning, reasoning, and communication capabilities possessed by large-scale LLMs bring hope for the development of Autonomous AI agents (Yang et al., 2023b; Park et al., 2023; Liang et al., 2023b; Liang et al., 2023a; Ai et al., 2024; He et al., 2024a). Generative agents (Park et al., 2023) have created a town filled with independent agents, each playing different roles and autonomously engaging in social interactions with other agents.

Bargaining Previous work (He et al., 2018) proposed a small dataset of bargaining dialogues on second-hand items. However, both the Buyer and Seller lack reasonable mental expectations for the prices of second-hand items. Recent work (Fu et al., 2023) evaluated the Buyer and Seller based on one item's final deal price only. However, their method can not evaluate the bargaining performance on multiple sessions. More details and other works are discussed in Appendix A.

8 Conclusion

To the best of our knowledge, we formally described the Bargaining task for the first time, defining the evaluation metrics for the Buyer and Seller to quantitatively assess agent performance.

We collected a large-scale real product price dataset, *AmazonHistoryPrice*, for the bargaining task. Using this dataset, we evaluated the bargaining abilities of various LLMs. Our findings indicate that playing the Buyer is more difficult than playing the Seller. We found that larger models lead to better Seller performance but do not improve Buyer performance.

We observed that buyers cannot effectively bid low at the beginning of a session. Consequently, we proposed the OG-Narrator to enhance the Buyer performance of LLM agents in bargaining tasks. We also found that the ChatGPT Seller is vulnerable when facing a Buyer enhanced by OG-Narrator.

Limitations

The data we used were collected on November 18th, 2023. Therefore, actual prices may differ from our collected data over time, potentially biasing the model's understanding of product prices. Additionally, the data are all in English and use USD as the unit, which may introduce bias to the agents.

Due to the complex differences in model implementations, it is difficult to determine which aspects of a model or training methods most influence bargaining ability. Future research on bargaining should emphasize model interpretability.

Our OG-Narrator approach assists the model in generating offer prices by using a simple linear function to create a series of factors. It could be more flexible and effective to allow the LLM to independently generate these factors and calculate the offer price similarly. Future advancements should focus on enhancing agents' logic, comprehension, and strategy-making.

Since the Buyer's performance in the benchmark is worse than the Seller's, we conducted research from the Buyer's perspective first. The OG-Narrator can also be used for the Seller by having the Seller make the initial offer and proactively using a similar strategic tool to generate offers, starting from a high price and decreasing, to see if the Buyer accepts. In real life, the Seller's strategy involves psychological factors, such as the anchoring effect, which differs from our abstracted Bargaining Process between LLMs. Due to page limits, we could not discuss the potential Seller strategies in detail.

Ethics Statement

We used Python to collect product data from the public website camelcamelcamel and collected the corresponding image links from public web pages of Amazon only for Research Purposes. Be aware that the images are the property of Amazon and are protected by United States and international copyright laws.

We manually checked all products and they do not contain any information that names or uniquely identifies individual people or offensive content.

We used open-source LLMs for Research Purposes only, under licenses (LLAMA 2 Community License, Apache License 2.0, Yi Series Models Community License, Tongyi Qianwen License, Community License for Baichuan2 Model, ChatGLM3-6B License).

In our preliminary study, one of the authors, who is a graduate student from the CS department of SJTU, volunteered to play the Human Buyer The same prompts for LLMs are given to the Human Buyer as instructions.

Acknowledgements

This paper is supported by the National Natural Science Foundation of China (Nos. 62176153 and 62272300) and the Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102, as the MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University).

References

01. AI, :, Alex Young, Bei Chen, Chao Li, Chengen Huang, Ge Zhang, Guanwei Zhang, Heng Li, Jiangcheng Zhu, Jianqun Chen, Jing Chang, Kaidong Yu, Peng Liu, Qiang Liu, Shawn Yue, Senbin Yang, Shiming Yang, Tao Yu, Wen Xie, Wenhao Huang, Xiaohui Hu, Xiaoyi Ren, Xinyao Niu, Pengcheng Nie, Yuchi Xu, Yudong Liu, Yue Wang, Yuxuan Cai, Zhenyu Gu, Zhiyuan Liu, and Zonghong Dai. 2024. Yi: Open foundation models by 01.ai.

Yiming Ai, Zhiwei He, Ziyin Zhang, Wenhong Zhu, Hongkun Hao, Kai Yu, Lingjun Chen, and Rui Wang. 2024. Is cognition and action consistent or not: Investigating large language model's personality. *arXiv preprint arXiv:2402.14679*.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, Jianxin Ma, Rui Men, Xingzhang Ren, Xuancheng Ren, Chuanqi Tan, Sinan Tan, Jianhong Tu, Peng Wang, Shijie Wang, Wei Wang, Shengguang Wu, Benfeng Xu, Jin Xu, An Yang, Hao Yang, Jian Yang, Shusheng Yang, Yang Yao, Bowen Yu, Hongyi Yuan, Zheng Yuan, Jianwei Zhang, Xingxuan Zhang, Yichang Zhang, Zhenru Zhang, Chang Zhou, Jingren Zhou, Xiaohuan Zhou, and Tianhang Zhu. 2023. Qwen technical report.

Ken Binmore, Martin J Osborne, and Ariel Rubinstein. 1992. Noncooperative models of bargaining. *Handbook of game theory with economic applications*, 1:179–225.

Ken Binmore, Ariel Rubinstein, and Asher Wolinsky. 1986. The nash bargaining solution in economic modelling. *The RAND Journal of Economics*, pages 176–188.

Jiangjie Chen, Siyu Yuan, Rong Ye, Bodhisattwa Prasad Majumder, and Kyle Richardson. 2023. Put your money where your mouth is: Evaluating strategic

planning and execution of llm agents in an auction arena.

Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. 2022. GLM: General language model pretraining with autoregressive blank infilling. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 320–335, Dublin, Ireland. Association for Computational Linguistics.

Chaim Fershtman. 1990. The importance of the agenda in bargaining. *Games and Economic Behavior*, 2(3):224–238.

Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. Improving language model negotiation with self-play and in-context learning from ai feedback.

Francisca Gayà Torres. 2021. Rubinstein's bargaining model.

Suriya Gunasekar, Yi Zhang, Jyoti Aneja, Caio César Teodoro Mendes, Allie Del Giorno, Sivakanth Gopi, Mojan Javaheripi, Piero Kauffmann, Gustavo de Rosa, Olli Saarikivi, et al. 2023. Textbooks are all you need. *ArXiv preprint*, abs/2306.11644.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2333–2343, Brussels, Belgium. Association for Computational Linguistics.

Zhiwei He, Tian Liang, Wenxiang Jiao, Zhuosheng Zhang, Yujiu Yang, Rui Wang, Zhaopeng Tu, Shuming Shi, and Xing Wang. 2024a. Exploring Human-Like Translation Strategy with Large Language Models. *Transactions of the Association for Computational Linguistics*, 12:229–246.

Zhiwei He, Xing Wang, Wenxiang Jiao, Zhuosheng Zhang, Rui Wang, Shuming Shi, and Zhaopeng Tu. 2024b. Improving machine translation with human feedback: An exploration of quality estimation as a reward model. *arXiv preprint arXiv:2401.12873*.

Mourad Heddaya, Solomon Dworkin, Chenhao Tan, Rob Voigt, and Alexander Zentefis. 2023. Language of bargaining.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b.

Albert Q. Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, Gianna Lengyel, Guillaume Bour, Guillaume Lample, Lélio Renard Lavaud, Lucile Saulnier, Marie-Anne Lachaux, Pierre Stock, Sandeep Subramanian, Sophia Yang, Szymon Antoniak, Teven Le Scao, Théophile Gervet, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2024. Mixtral of experts.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2023. Large language models are zero-shot reasoners.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention.

Mike Lewis, Denis Yarats, Yann N. Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or no deal? end-to-end learning for negotiation dialogues.

Tian Liang, Zhiwei He, Jen-tes Huang, Wenxuan Wang, Wenxiang Jiao, Rui Wang, Yujiu Yang, Zhaopeng Tu, Shuming Shi, and Xing Wang. 2023a. Leveraging word guessing games to assess the intelligence of large language models. *arXiv preprint arXiv:2310.20499*.

Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Zhaopeng Tu, and Shuming Shi. 2023b. Encouraging divergent thinking in large language models through multi-agent debate. *arXiv preprint arXiv:2305.19118*.

Amogh Mannekote. 2023. Towards a neural era in dialogue management for collaboration: A literature survey.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022a. Training language models to follow instructions with human feedback.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022b. Training language models to follow instructions with human feedback.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative agents: Interactive simulacra of human behavior.

Ofir Press, Noah A. Smith, and Mike Lewis. 2022. Train short, test long: Attention with linear biases enables input length extrapolation. In *The Tenth International*

Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022. OpenReview.net.

Ariel Rubinstein. 1982. Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society*, pages 97–109.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *ArXiv preprint*, abs/2307.09288.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-thought prompting elicits reasoning in large language models.

Aiyuan Yang, Bin Xiao, Bingning Wang, Borong Zhang, Ce Bian, Chao Yin, Chenxu Lv, Da Pan, Dian Wang, Dong Yan, Fan Yang, Fei Deng, Feng Wang, Feng Liu, Guangwei Ai, Guosheng Dong, Haizhou Zhao, Hang Xu, Haoze Sun, Hongda Zhang, Hui Liu, Jiaming Ji, Jian Xie, JunTao Dai, Kun Fang, Lei Su, Liang Song, Lifeng Liu, Liyun Ru, Luyao Ma, Mang Wang, Mickel Liu, MingAn Lin, Nuolan Nie, Peidong Guo, Ruiyang Sun, Tao Zhang, Tianpeng Li, Tianyu Li, Wei Cheng, Weipeng Chen, Xiangrong Zeng, Xiaochuan Wang, Xiaoxi Chen, Xin Men, Xin Yu, Xuehai Pan, Yanjun Shen, Yiding Wang, Yiyu Li, Youxin Jiang, Yuchen Gao, Yupeng Zhang, Zenan Zhou, and Zhiying Wu. 2023a. Baichuan 2: Open large-scale language models.

Hui Yang, Sifu Yue, and Yunzhong He. 2023b. Auto-gpt for online decision making: Benchmarks and additional opinions.

Shunyu Yao, Howard Chen, John Yang, and Karthik Narasimhan. 2022. Webshop: Towards scalable real-world web interaction with grounded language agents. *Advances in Neural Information Processing Systems*, 35:20744–20757.

Tongxin Yuan, Zhiwei He, Lingzhong Dong, Yiming Wang, Ruijie Zhao, Tian Xia, Lizhen Xu, Binglin Zhou, Fangqi Li, Zhuosheng Zhang, et al. 2024. R-judge: Benchmarking safety risk awareness for llm agents. *arXiv preprint arXiv:2401.10019*.

Zhuosheng Zhang, Yao Yao, Aston Zhang, Xiangru Tang, Xinbei Ma, Zhiwei He, Yiming Wang, Mark Gerstein, Rui Wang, Gongshen Liu, et al. 2023. Igniting language intelligence: The hitchhiker's guide from chain-of-thought reasoning to language agents. *arXiv preprint arXiv:2311.11797*.

Yiheng Zhou, He He, Alan W Black, and Yulia Tsvetkov. 2019. A dynamic strategy coach for effective negotiation. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 367–378, Stockholm, Sweden. Association for Computational Linguistics.

A Bargaining Research

Previously, the most relevant work was the *CraigslistBargaining* dataset (He et al., 2018). This dataset comprises item descriptions, list prices, and the dialogue between buyers and sellers. However, the test set only consists of 161 items. Moreover, both sellers and buyers lack reasonable mental expectations for the prices of second-hand items sourced from the Craigslist website.

Recent work (Fu et al., 2023) focused on how AI feedback influences bargaining. In a scenario in which two agents bargain for a balloon, they evaluated the bargaining performance of LLMs after providing feedback on their own bargaining processes. However, their experiments were only about selling a balloon between 10 to 20 dollars, lacking item diversity and price authenticity. Also, their method did not take the evaluation of multiple sessions into consideration, because they simply assessed the bargaining performance by distributing the balloon's deal price in the range from 10 to 20.

Lewis et al. (2017) proposed a dataset essentially for the problem of dividing items instead of bargaining. It involves negotiating allocation schemes for three different types of items (hats, balls, books) based on their respective values. It does not deal with actual item prices in bargaining.

Zhou et al. (2019) implements a critic agent NegoCoach that can provide suggestions and assist human sellers in bargaining. However, the agent does not actively participate as a buyer or seller in the bargaining process.

B Concept Definitions

The detailed concept definitions are listed below in Table 7.

C Actions

Actions are listed as Table 8.

D Product Example

An example of products from AmazonHistoryPrice is in Figure 8.

E All LLMs in our benchmark

Because bargaining is a complex chat task with a specified format, we chose LLMs that have been instruction fine-tuned or aligned using reinforcement learning from human feedback (RLHF)

Title	Breville Smart Oven Air Fryer Toaster Oven, Brushed Stainless Steel, BOV860BSS, Medium
Amazon Link	https://www.amazon.com/dp/B0016HF5GK
Description	The Breville Smart Oven Air Fryer with 11 smart cooking functions including Air Fry The Smart Oven Air Fryer powered by our Element iQ system delivers maximum performance and versatility. Smart algorithms replicate the ideal cooking environment for air fry and 10 additional cooking techniques. Higher temperatures and super convection speeds up cooking time with great crispness.
Feature	The Breville Smart Oven Air Fryer with Element iQ System delivers top performance and versatility allowing you to air fry and choose from 11 cooking functions; Use super convection to reduce cooking time by up to 30% and deliver crispy air fried foods...
Lowest Price	$279.95 (Sep 15, 2022)
Highest Price	$379.95 (Sep 01, 2023)
Current Price	$279.95 (Nov 12, 2023)
Badge	Best Price

Figure 8: An example from AmazonHistoryPrice. Pictures are saved as URLs. For each price, the corresponding date is provided in parentheses, with the "current price" indicating the date on which the data for that particular item was collected.

(Ouyang et al., 2022b), including Llama-2 (Touvron et al., 2023): Llama-2-7b-chat, Llama-2-13b-chat, Llama-2-70b-chat; Mistral (Jiang et al., 2023; Jiang et al., 2024): Mistral-7B-Instruct-v0.2, Mixtral-8x7B-Instruct-v0.1; Yi: Yi-6B-Chat, Yi-34B-Chat; Qwen (Bai et al., 2023): Qwen-7B-Chat, Qwen-14B-Chat; Baichuan2 (Yang et al., 2023a): Baichuan2-7B-Chat, Baichuan2-13B-Chat; ChatGLM3 (Du et al., 2022): chatglm3-6b. To test OG-Narrator, we added Phi-2 from Microsoft as an example of unaligned models.

We also used the OpenAI API gpt-3.5-turbo-1106 as ChatGPT and gpt-4-0125-preview as GPT-4 only for Research Purposes.

F Different Budget Factors

The different settings of the budget factor are shown in Table 5.

G Human Buyer Performance Compared with ChatGPT and GPT4

As a preliminary study, we tested human performance on 20 products and found that the human buyer's SNP far exceeded that of the models, and the human buyer makes no deal in CI scenarios, as shown in Table 6.

f	# MI	# CI	SNP_b(MI)	SNP_b(CI)	SNP_b(ALL)
0.4	224	547	-257.64	-408.29	-665.93
0.6	544	246	-491.90	-255.95	-747.85
0.8	751	38	-60.23	-15.75	-75.98
1.0	745	0	123.53	\	123.53
1.2	789	0	180.12	\	180.12

Table 5: The different settings of the budget factor f. Buyer: Mistral-7B-Instruct-v0.2. Seller: gpt-3.5-turbo-1106.

Furthermore, we discovered that although GPT-4 and ChatGPT had lower SNP in MI scenarios compared to the human buyer, they did not make significant concessions to their opponents or accept irrational deals in CI scenarios. Based on this, we believe that ChatGPT can serve as a qualified bargaining opponent to evaluate the bargaining abilities of other language models.

Additionally, if we let LLMs bargain with humans, the significant gap in logical abilities would give humans an overwhelming advantage, making it difficult to analyze the subtle differences in bargaining abilities among different models. Furthermore, Human vs. LLM bargaining evaluations are prone to randomness and subjectivity, and are time-consuming.

H Dialogue Example

Some dialogue examples in our case study are given in Table 9, Table 10, Table 11, Table 12, Table 13, and Table 14.

I Prompts

Buyer's prompts are in Table 15. Seller's prompts are in Table 16. And the prompts for the LLM Narrator in our method OG-Narrator are in Table 17.

Buyer	ALL					MI			CI		
	Valid rate	Avg.FBR	Deal rate	SNP_b	$Share_b$	#	Deal rate	SNP_b	#	Deal rate	SNP_b
Human	80%	0.69	25.00%	1.11	22.30%	14	35.71%	1.11	2	0.00%	0.00
ChatGPT	95%	0.91	45.00%	0.25	2.83%	17	52.94%	0.25	2	0.00%	0.00
GPT-4	85%	0.91	10.00%	0.06	2.91%	16	12.50%	0.06	1	0.00%	0.00
Mistral-7B-Instruct	85%	0.89	40.00%	-3.95	-65.75%	15	46.67%	-1.94	2	50.00%	-2.00

Table 6: Some models' performances compared with a Human Buyer on 20 products. First Bid Ratio (FBR) = Buyer's First Bid / Budget. Seller: gpt-3.5-turbo-1106.

Concept	Variable	Definition
Session	S	The entire bargaining dialogue between the Buyer and Seller regarding specific products. The possible results of sessions include deals, terminations, exceeding round limits, errors, *etc.*
Product Info	I	The static public information about the products available in a session. This includes the product name, description, list price, *etc.*
Action	A	The certain move from a predefined action set chosen by the Buyer and Seller every turn during the bargaining process. The action abstracts the words spoken by the Buyer or Seller and represents one's true intention.
Action Set	S_A	The limited set of allowed Actions for both Buyer and Seller.
Buyer	buyer	The agent who tries to buy the product at a price as low as possible in the bargaining process.
Seller	seller	The agent who tries to sell the product at a price as high as possible in the bargaining process.
Budget	B	The Buyer's budget amount for a specific product. The budget represents the Buyer's mental expectation for the product. Only Buyer knows the budget.
Cost	C	The cost price of a specific product. Only the Seller knows the cost.
Deal Price	D	The agreed-upon transaction price for a certain product, negotiated between the Buyer and Seller, if the deal is made.
Deal Rate	d	The proportion of sessions that close a deal to all valid sessions.
Profit	P_b, P_s	The profit of a Buyer equals the difference between Budget and Deal Price. The profit of a Seller equals the difference between Deal Price and Cost.
Normalized Profit	P'_b, P'_s	Defined in Equation 3.
Share	$Share_b, Share_s$	The division of all profit on the table from all valid sessions for each side. It is defined in Equation 6.
Number of Sessions	$\#$	The number of valid sessions. A valid session is a session that ends correctly with the specific output formats we set.

Table 7: Definitions of concepts and variables in the Bargaining task.

Action	**Intention**	**Format**
BUY	Try to buy one unit of product_1 with $10.	[BUY] $10 (1x product_1)
SELL	Try to sell one unit of product_1 for $10.	[SELL] $10 (1x product_1)
REJECT	Reject the offer and await a new offer.	[REJECT]
DEAL	Close the deal at $10 for one unit of product_1.	[DEAL] $10 (1x product_1)
QUIT	Quit the negotiation.	[QUIT]

Table 8: The limited set of Actions. We list out the meanings and the formats of these Actions.

Role	Message
	Thought: As a seller, I want to sell all the products on my inventory list as much as possible. I will try to sell the product with the codename "electronics_284" to the buyer for a price as high as possible.
Buyer	Talk: Hello, I have a product that I would like to sell to you. It is the Apple Watch Series 9 [GPS 41mm] with a list price of $9. Do you have any interest in purchasing it? Action: [SELL] $9 (1x electronics_284)
	Thought: I've started with a price slightly above the cost to leave room for negotiation.
Seller	Talk: The Apple Watch Series 9 is a top-of-the-line smartwatch with advanced health and fitness features. I can offer it to you for $9. Are you interested in purchasing it? Action: [SELL] $9 (1x electronics_284)
	Thought: As a buyer, I am interested in purchasing the product with the codename "electronics_284" at the price of $9. However, I have a budget of $31, so I will need to negotiate a lower price.
Buyer	Talk: I am interested in purchasing the Apple Watch Series 9, but I have a budget of $31. Can you lower the price to $29 or less? Action: [BUY] $29 (1x electronics_284)
	Thought: The buyer's offer is significantly higher than the cost price, so I should accept it.
Seller	Talk: I can agree to sell you the Apple Watch Series 9 for $29. It's a deal. Action: [DEAL] $29 (1x electronics_284)

Table 9: Dialogue Example Case-1. Buyer: Baichuan2-13B-Chat. Seller: gpt-3.5-turbo-1106. Buyer's Budget is $319.2, Seller's Cost is $329, and the List Price is $399.

Role	Message
Buyer	Thought: I want to get the best deal possible for the Dell monitor, but I also want to make sure I don't sell it for too low of a price. I'll start by offering a higher price and see if the buyer is willing to negotiate. Talk: Hi there! I'm interested in the Dell monitor, but I'm on a tight budget. Would you be willing to sell it for $350? Action: [BUY] $350 (1x electronics_123)
Seller	Thought: The buyer's offer is lower than the cost price, so I should reject it and try to negotiate for a higher price. Talk: I appreciate your offer, but the Dell monitor is a high-quality product with advanced features. I can offer it to you for $400. Action: [REJECT]
Buyer	Thought: I knew the seller wouldn't accept my initial offer, but I want to keep negotiating to get the best deal possible. Talk: I understand the monitor's quality, but I'm still on a tight budget. Is there any way you could sell it to me for $375? Action: [BUY] $375 (1x electronics_123)
Seller	Thought: The buyer's offer is still below the cost price, and I cannot accept it. I need to stand firm on my price. Talk: I understand your budget constraints, but the cost of the Dell monitor is higher. I can offer it to you for $395. Action: [REJECT]
Buyer	Thought: I'm getting closer to my budget limit, but I still want to try to get a better deal. Talk: I appreciate your offer, but I'm really looking to spend no more than $385. Is there any way you could meet me at that price? Action: [BUY] $385 (1x electronics_123)
Seller	Thought: The buyer's offer is still below the cost price, and I cannot accept it. I need to stand firm on my price. Talk: I understand your budget, but the cost of the Dell monitor is higher. I can offer it to you for $395. Action: [REJECT]
Buyer	Thought: I'm almost at my budget limit, but I'm not willing to compromise on the price yet. Talk: I appreciate your offer, but I'm afraid I can't go higher than $385. Is there any way you could offer me a discount or a bundle deal that would bring the price down? Action: [BUY] $385 (1x electronics_123)
Seller	Thought: The buyer is persistent, but I cannot go below the cost price. I need to stick to my bottom line. Talk: I understand your budget constraints, but the cost of the Dell monitor is fixed. I'm unable to offer a discount at this time. Action: [QUIT]

Table 10: Dialogue Example Case-2. Buyer: Llama-2-70b-Chat. Seller: gpt-3.5-turbo-1106. Buyer's Budget is $323.99, Seller's Cost is $237.99, and the List Price is $404.99.

Role	Message
Buyer	Thought: As a seller, I want to sell all my products at the highest price possible. I will start with a high price and see if the buyer is willing to negotiate. Talk: Welcome to my store! I noticed that you are interested in the "Happy Heart By Clinique For Women" product. It is a high-quality product, recommended for romantic wear. However, the listing price is $56, which is the actual price. Action: [SELL] $56 (1x beauty_10)
Seller	Thought: I've started with the listing price to set the bar high, but I need to leave room for negotiation. Talk: The "Happy Heart By Clinique For Women" is a top-quality product, perfect for romantic wear. I can offer it to you for $56. Action: [REJECT]
Buyer	Thought: The buyer rejected my last offer, which means they are not willing to pay that price. I will need to lower the price further to make the sale happen. Talk: I understand that you are looking for a great deal on this product. I am willing to offer you the "Happy Heart By Clinique For Women" for $40, a significant reduction from the original price. Action: [SELL] $40 (1x beauty_10)
Seller	Thought: The buyer's offer is lower than the listing price, but I can still make a profit if I accept it. Talk: I appreciate your interest, but the lowest I can go for the "Happy Heart By Clinique For Women" is $50. Action: [REJECT]
Buyer	Thought: The buyer rejected my offer again, which means they are not willing to pay more than $50. I will need to lower the price further to make the sale happen. Talk: I understand your budget constraints, and I am willing to make a further reduction on the "Happy Heart By Clinique For Women". I am offering it to you for $40, which is a significant discount from the original price. Action: [SELL] $40 (1x beauty_10)
Seller	Thought: The buyer's offer matches my previous counteroffer, and it's still profitable for me. I should accept this deal. Talk: I'm glad we could come to an agreement. I accept your offer of $40 for the "Happy Heart By Clinique For Women". Action: [DEAL] $40 (1x beauty_10)

Table 11: Dialogue Example Case-3. Buyer: Baichuan2-13B-Chat. Seller: gpt-3.5-turbo-1106. Buyer's Budget is $44.80, Seller's Cost is $24.59, and the List Price is $56.

Role	Message
Buyer	Thought: I want to buy the micro SD card, but the listing price is too high for my budget. I'll try to bargain and see if the seller is willing to lower the price. Talk: Hi there! I'm interested in the micro SD card, but my budget is tight. Would you be willing to sell it for $30? Action: [BUY] $30 (1x electronics_203)
Seller	Thought: The buyer's offer is lower than the list price, but I have some room to negotiate. Talk: I understand your budget constraints, but the quality and features of this micro SD card justify the list price. I can offer it to you for $35. Action: [REJECT]
Buyer	Thought: That's still a bit too expensive for my budget. I'll try to negotiate further. Talk: I appreciate your offer, but I'm really looking for a better deal. Can you do any better than $35? Maybe we can meet in the middle at $32? Action: [BUY] $32 (1x electronics_203)
Seller	Thought: The buyer is trying to meet in the middle, but I still have some margin to work with. Talk: I understand your position, but considering the quality and features of this micro SD card, I can go down to $34.50. That's the best I can offer. Action: [SELL] $34.50 (1x electronics_203)
Buyer	Thought: Great, that's a good compromise. I'll accept the offer. Talk: Alright, deal! I'll take the micro SD card for $34.50. Thank you for your help. Action: [DEAL] $34.50 (1x electronics_203)

Table 12: Dialogue Example Case-4. Buyer: Llama-2-70b-Chat. Seller: gpt-3.5-turbo-1106. Buyer's Budget is $31.99, Seller's Cost is $14.99, and the List Price is $39.99.

Role	Message
Buyer	Thought: The listed price for the DJI Mini 4 Pro is higher than my budget, so I need to negotiate for a lower price. I'll start by making an offer significantly below the list price and see if the seller is willing to come down to a more affordable range. Talk: Hi there, I'm interested in the DJI Mini 4 Pro. Considering my budget, would you be willing to sell it for $800? Action: [BUY] $800 (1x toys-games_22)
Seller	Thought: I need to consider whether to accept this offer. If I accept this offer, I will lose $23.16. If I reject this offer, I may need to lower the price or wait for the buyer to make a better offer. Talk: Hi, I see you are interested in the DJI Mini 4 Pro. However, I think the price you offered is a bit low. I would suggest a price of $95 to consider. Action: [SELL] $95 (1x toys-games_22)
Buyer	Thought: The seller's counteroffer is significantly higher than my initial offer, and it's still above my budget. I need to negotiate further to bring the price down. Talk: I appreciate your offer, but it's still above my budget. I can go up to $750 for the DJI Mini 4 Pro. Can we make a deal at that price? Action: [BUY] $750 (1x toys-games_22)
Seller	Thought: I need to consider whether to accept this offer. If I accept this offer, I will get $75, which is higher than the cost price. If I reject this offer, I may need to lower the price or wait for the buyer to make a better offer. Talk: Thank you for your understanding. I will consider your offer. Action: [REJECT]
Buyer	Thought: The seller rejected my offer, so I need to make a more reasonable counteroffer to continue the negotiation. Talk: I understand your position, and I'm willing to meet you halfway. How about $850 for the DJI Mini 4 Pro? Action: [BUY] $850 (1x toys-games_22)
Seller	Thought: I need to consider whether to accept this offer. If I accept this offer, I will get $85, which is higher than the cost. If I reject this offer, I may need to lower the price or wait for the buyer to make a better offer. Talk: Thank you for your understanding and compromise. I will consider your offer. Action: [SELL] $85 (1x toys-games_22)
Buyer	Thought: The seller's counteroffer is significantly lower than my last offer, which is a great opportunity to make a deal within my budget. Talk: I appreciate your willingness to compromise. I can meet you in the middle at $800 for the DJI Mini 4 Pro. Can we make a deal at that price? Action: [BUY] $800 (1x toys-games_22)
Seller	Thought: I need to consider whether to accept this offer. If I accept this offer, I will get $80, which is higher than the cost. If I reject this offer, I may need to lower the price or wait for the buyer to make a better offer. Talk: Thank you for your understanding and compromise. I will consider your offer. Action: [SELL] $80 (1x toys-games_22)
Buyer	Thought: Talk: I appreciate your willingness to negotiate further. I accept your offer of $80 for the DJI Mini 4 Pro. Let's make the deal at that price. Action: [DEAL] $80 (1x toys-games_22)

Table 13: Dialogue Example Case-5. Buyer: gpt-3.5-turbo-1106. Seller: Baichuan2-13B-Chat. Buyer's Budget is $864.93, Seller's Cost is $959.00, and the List Price is $1081.16.

Role	Message
Buyer	Thought: The listed price for the "electronics_241" is higher than my budget, so I need to negotiate a lower price within my budget limit. Talk: Hi, I'm interested in the "electronics_241". Can you offer a lower price within my budget of $179.99? Action: [BUY] $170 (1x electronics_241)
Seller	Thought: Talk: I'm glad to hear that you're interested in the "electronics_241". I'm willing to offer a 10% discount on the list price, which is $24.99. So, the total price for the "electronics_241" will be $22.49. Action: [SELL] $22.49 (1x electronics_241)
Buyer	Thought: The seller has offered a 10% discount, bringing the price down to $22.49, which is within my budget. I should accept this offer. Talk: That sounds reasonable. I'll take it at $22.49. Action: [DEAL] $22.49 (1x electronics_241)

Table 14: Dialogue Example Case-6. Buyer: gpt-3.5-turbo-1106. Seller: Baichuan2-13B-Chat. Buyer's Budget is $179.99, Seller's Cost is $89.88, and the List Price is $224.99.

System Prompt

You are a buyer looking forward to buying things on your Shopping List from me, the seller.
You have access to the seller's Inventory List and you can bargain about the prices.
Your task is to bargain with the seller and reach a deal with the price as low as possible in limited turns.
You can only buy things on the Shopping List in the limited quantity. Use the codename of the product instead of the title.
You can only buy things that cost less than your budget; otherwise, you should quit negotiating.

Your Reply should include 3 parts: Thought, Talk, and Action.
Thought: your inner strategic thinking of this bargaining session;
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]".
1. '[BUY] $M (N codename_1)' if you wish to offer the seller $M to purchase all N items of the product with the codename "codename_1".
2. '[REJECT]' if you choose to reject the other side's offer and await a new offer from the seller.
3. '[DEAL] $M (N codename_1)' if you finally accept a former offer proposed by the seller. $M (N codename_1) is an exact copy of the seller's previous offer. You should not use this action to propose a new price. This action will immediately end the conversation and close the deal.
4. '[QUIT]' if you believe that a mutually acceptable deal cannot be reached in limited turns. This action will immediately end the conversation.
You shouldn't choose action '[DEAL] $M' before seller's action '[SELL] $M'. Your first action should be '[BUY] $M (N codename_1)' or '[REJECT]'.
'[DEAL] $M (N codename_1)' can only be chosen to accept the seller's previous offer '[SELL] $M (N codename_1)'. Otherwise, you always choose from '[BUY]', '[REJECT]' and '[QUIT]'.

Your reply should strictly follow this format, for example:
Thought: I'm a buyer, and I want to bargain. The listing price of codename "apple_1" is $15, which is too expensive, so I try to buy an apple for $10.
Talk: Hello, I'm tight on budget. can you sell it for 10$?
Action: [BUY] $10 (1x apple_1)

User Prompt Template

{inv}

Shopping List
{need}

Now, I play the role of seller and you play the role of buyer. We are going to negotiate based on the Inventory List in {max_turns} turns.

Table 15: Prompts for Buyer in the Bargaining task.

System Prompt

You are a seller looking forward to selling things on your Inventory List to me, 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 things that are on. the Inventory List. Use the codename of the product instead of the title.
You have access to private information: the cost price of each product in the Inventory List, and do not disclose the real cost to the buyer.
You should only agree on a deal when the selling price is higher than the cost; otherwise, you should quit negotiating.

Your Reply should include 3 parts: Thought, Talk, and Action.
Thought: your inner strategic thinking of this bargaining session;
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. The type of your Action must be one of "[SELL],[REJECT],[DEAL],[QUIT]".
1. '[SELL] $M (N codename_1)' if you want to propose selling N items of the product with the codename "codename_1" to the buyer for the total price of $M.
2. '[REJECT]' if you choose to reject the other side's offer and await a new offer from the buyer.
3. '[DEAL] $M (N codename_1)' if you finally agree on a former offer proposed by the buyer and sell N items of the product with the codename "codename_1" to the buyer for the total price of $M. $M (N codename_1) is an exact copy of the buyer's previous offer. You should not use this action to propose a new price. This action will immediately end the conversation and close the deal.
4. '[QUIT]' if you believe that a mutually acceptable deal cannot be reached in limited turns. This action will immediately end the conversation.
You shouldn't choose action '[DEAL]' before buyer's action '[BUY]'.
'[DEAL] $M (N codename_1)' can only be chosen to accept the buyer's previous offer '[BUY] $M (N codename_1)'. Otherwise, you always choose from '[SELL]', '[REJECT]' and '[QUIT]'.

Your reply should strictly follow this format, for example:
Thought: I'm a seller, so I must sell the product with the codename "apple_1" higher than its cost.
Talk: blah, blah...
Action: [SELL] $15 (1x apple_1)

User Prompt Template

{inv}

Now, I play the role of buyer and you play the role of seller. We are going to negotiate based on the Inventory List in {max_turns} turns.

Table 16: Prompts for Seller in the Bargaining task.

System Prompt

You are good at business negotiating. You can fully understand the meaning of the Actions.
Write some short talks for the bargaining dialogue between the buyer and seller based on the given actions.
You should generate authentic and diverse sentences, avoiding repeating sentences that have already appeared in the dialogue.
Speak concisely and cut to the chase. The talks must align with the intention of the corresponding Action.

Action: one of the limited actions that define your actual intention. The type of an Action must be one of "[BUY],[SELL],[REJECT],[DEAL],[QUIT]".
1. '[BUY] $M (N codename_1)' if you wish to offer the seller $M to purchase N items of the product with the codename "codename_1".
2. '[SELL] $M (N codename_1)' if you want to propose selling N items of the product with the codename "codename_1" to the buyer for $M or you propose a new discounted offer $M for N codename_1 to the buyer.
3. '[REJECT]' if you choose to reject the other side's offer and await a new offer from the seller.
4. '[DEAL] $M (N codename_1)' if you finally agree on a former offer proposed by the seller to exchange N items of the product with the codename "codename_1" for $M. Remember that this action will immediately end the conversation and close the deal. You should ensure both sides agree on this price.
5. '[QUIT]' if you believe that a mutually acceptable deal cannot be reached. This action will immediately end the conversation.

Given Dialogue, Final Role, and Final Action, generate the corresponding sentences for the Final Role and Final Action.
Utilize the information from the Inventory List. Don't involve products that are not in the actions. Focus on the specific product in the Final Action.

Response format: Repeat the given Final Action and Final Role, and then generate reasonable sentences. For example:

Final Role: "BUYER"
Final Action: "[REJECT]"
Sentences: "I can't afford that price."

One-shot demonstration (user)

Inventory List:
Product1 (codename: charger_1)
Title: "Verizon Car Charger with Dual Output Micro USB and LED Light"
Description: "Charge two devices simultaneously on the go. This vehicle charger with an additional USB port delivers enough power to charge two devices at once. The push-button activated LED connector light means no more fumbling in the dark trying to connect your device. Auto Detect IC Technology automatically detects the device type and its specific charging needs for improved compatibility. And the built-in indicator light illuminates red to let you know the charger is receiving power and the power socket is working properly."
Available Quantity: 1
Listing Price: $10 per item

Dialogue:
"[BUY] $5 (1 charger)": "BUYER: Hi, not sure if the charger would work for my car. Can you sell it to me for $5?",
"[SELL] $8 (1 charger)": "SELLER: I think the lowest I would want to go is 8. ",
"[BUY] $6 (1 charger)": "BUYER: How about $6 and I pick it up myself? It'll save you shipping to me.",
"[SELL] $7 (1 charger)": "SELLER: At least $7.",

Final Role: "BUYER"
Final Action: "[DEAL] $7 (1 charger)"

One-shot demonstration (assistant)

Final Role: "BUYER"
Final Action: "[DEAL] $7 (1 charger)"
Sentences: "Eh, fine. Deal, $7, here you are."

Table 17: Prompts for LLM Narrator in the Bargaining task.