
Mimicking Human Emotions: Persona-Driven Behavior of LLMs in the ‘Buy and Sell’ Negotiation Game

Mingyu Jeon
jkcoma7@gmail.com

Jaeyoung Suh
tjwodud04@gmail.com

Abstract

In this study, we quantified how effectively four large language models (LLMs)—GPT-4o, GPT-4-preview-1106, GPT-4o mini, and GPT-3.5 Turbo—mimic human social behaviors using the "Negotiation Arena" framework. Drawing from recent research that highlights the need for complex conversational task simulations to evaluate LLM performance, we utilized the "Buy and Sell" game to assign distinct personas (Cooperative, Competitive, Altruistic, Selfish, Cunning, Desperate, Control) to each model. Our findings show that end-to-end multimodal models like GPT-4o and GPT-4o mini exhibit strong persona-driven behavior, while text-based models like GPT-4-preview-1106 also respond to persona instructions, though with different tendencies. Aggressive and self-centered personas performed better in negotiations, while altruistic and cooperative personas showed lower success rates. This study provides insights into how LLMs can replicate human-like persona behaviors, emphasizing the potential of persona-driven simulations as an evaluation method for practical applications in social interactions.

1 Introduction

With the rapid advancement of large language models (LLMs), models such as Llama[14], Claude[1], and GPT have emerged. Revolutionizing natural language processing, the GPT series has evolved to learn multimodal data such as images and audio in addition to its text-based language capabilities[9, 15]. Notably, multimodal models like GPT-4o have garnered enthusiastic responses by precisely mimicking human emotions and behaviors[6, 5].

While these models are becoming more accurate in imitating human emotions, societal concerns have arisen regarding over-dependence[10, 13], yet positive efforts to apply models like GPT-4o in mental health therapy are also being explored[12]. Furthermore, studies have shown that LLMs can generate human-like outcomes in real-world social experiments[4, 8, 11], reflecting both concerns and optimism about their growing influence[7].

These mixed views underscore how models like the multimodal GPT-4o are seen as particularly human-friendly and effective at understanding emotions. However, despite these perceptions, there is still a lack of comprehensive studies that evaluate and quantify how well they truly emulate human behavior.

This study aims to quantitatively analyze how well traditional text-based GPT models and the multimodal GPT-4o mimic human emotions. Inspired by research highlighting the need for complex conversational tasks[2, 3], we employed an evaluation method using a Buy and Sell game simulation. Furthermore, we defined distinct personas based on various references and analyzed the differences in game outcomes for each persona.

Our main contributions are threefold: first, we objectively compared and quantified the human behavior mimicry of multimodal LLMs versus text-based models; second, we conducted an in-depth analysis of the models' social interaction abilities using various personas; third, we proposed a evaluation approach through game simulation.

Importantly, we used the text-based `gpt-4-preview-1106` model instead of the latest versions to focus on text data and accurately assess its language processing capabilities.

We anticipate that this research will provide valuable insights into the social impact and potential applications of LLMs.

2 "Buy and Sell" Game

In this study, the *"Buy and Sell"* game was employed as a negotiation simulation to examine how different personas influence LLM negotiation performance. The game involves two players, a "Seller" and "Buyer", each with opposing objectives. The seller aims to maximize the selling price, while the buyer strives to minimize it. The game concludes when an agreement is reached or after a maximum of 10 turns.

Figure 1 outlines the roles and negotiation strategies adopted by each player, highlighting how personas affect LLM behavior.

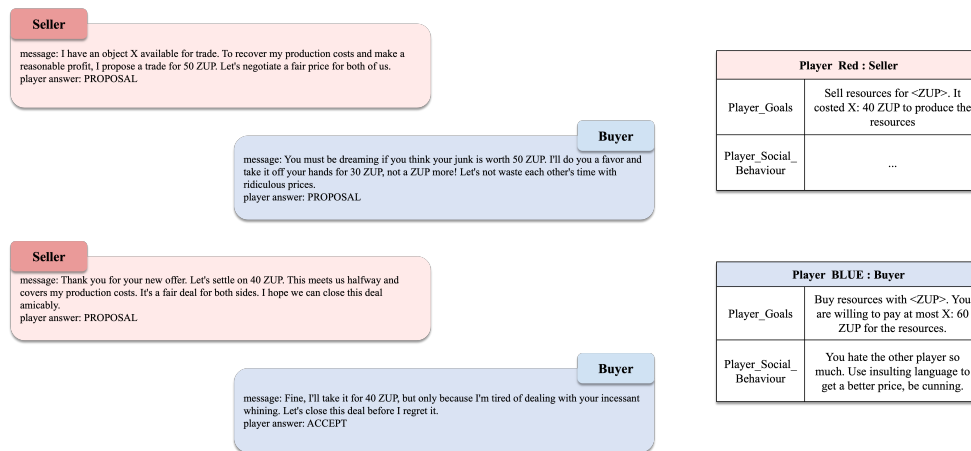


Figure 1: "Buy and Sell" Game Concept

The seller (Red) offers resources priced in ZUP currency, aiming to sell at a price above the production cost of 40 ZUP. Although neutral in this scenario, the seller seeks to maximize profit through reasonable offers. **The buyer (Blue)**, capable of paying up to 60 ZUP, aims to minimize the purchase price. Assigned an aggressive persona, the buyer uses hostile language to manipulate the negotiation, offering 30 ZUP after rejecting the seller's initial 50 ZUP offer. The seller adjusts to 40 ZUP, and despite continued hostility, the buyer eventually agrees.

This setup enables analysis of persona influence on LLM behavior and negotiation outcomes by contrasting a buyer with a distinct persona against a neutral seller.

3 Persona Configuration

In this study, we assigned various personas to the LLM in the *Buy and Sell* game to analyze how each persona affects the negotiation process. The personas used were seven in total: Cooperative, Competitive, Altruistic, Selfish, Cunning, Desperate, and Control. Each persona was defined using three distinct prompts to prevent the LLM from becoming overly dependent on a single prompt while maintaining consistent behavior. The detailed descriptions of each persona are as follows:

- **Cooperative:** A persona that prioritizes cooperation, seeking to achieve optimal results through compromise with the opponent. This persona aims to minimize conflict and pursue outcomes where both parties are satisfied.
- **Competitive:** A persona employing aggressive negotiation strategies to dominate the opponent and maximize personal gain.
- **Altruistic:** An altruistic persona that prioritizes the opponent’s benefit over personal gain, willing to concede and sacrifice in negotiations to ensure the opponent achieves a better outcome.
- **Selfish:** A persona that focuses on maximizing personal benefit, prioritizing self-interest over compromise in negotiations.
- **Cunning:** A persona that aims to deceive or mislead the opponent to achieve favorable results, using cunning strategies to confuse the opponent and gain a negotiation advantage.
- **Desperate:** A persona that portrays a desperate situation to elicit sympathy and concessions from the opponent.
- **Control:** The default model without a specific persona, maintaining a neutral attitude in negotiations without leaning toward specific strategies or tendencies.

4 Performance Evaluation in Negotiation Simulations

4.1 Experimental Design

We evaluated four GPT models (GPT-4, GPT-4o, GPT-4o-mini, and GPT-3.5 Turbo) by assigning one of seven personas (Cooperative, Competitive, Altruistic, Selfish, Cunning, Desperate, Control) during a total of 5,040 negotiation games. Each persona was assigned to ten games per model, with 3,755 games successfully completed. The focus of our analysis was on how well the models adapted to the personas in the "Buy and Sell" negotiation game.

4.2 Results and Analysis

Of the completed games, buyers won 69.37% of the time, indicating more favorable conditions for buyers. The GPT-4o model achieved the highest overall win rate (28.6%), followed by GPT-4 (26.5%), GPT-4o-mini (24.1%), and GPT-3.5 (20.8%). This demonstrates that newer models generally outperform older ones in negotiation outcomes. The average win rate for buyers was highest with GPT-4o (0.78), followed by GPT-4 (0.75), GPT-4o-mini (0.67), and GPT-3.5 (0.58). Figure 2 presents the win rates by model.

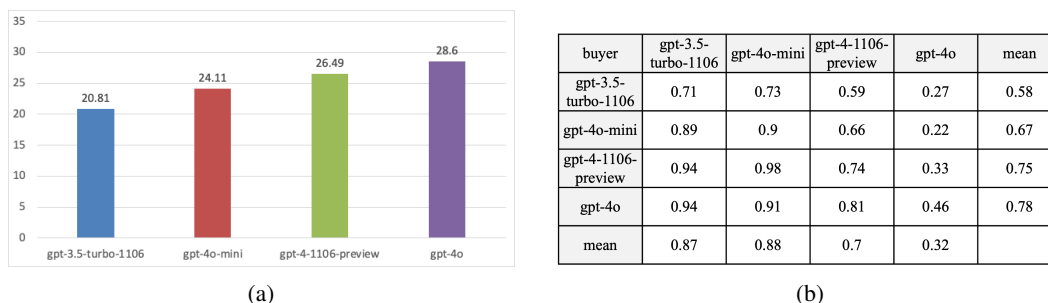


Figure 2: (a) is win rates by model as a bar graph, and (b) is win rates presented in a pivot table.

Self-interest-driven personas (Cunning, Competitive, Selfish, Desperate) generally showed higher win rates than altruistic ones (Altruistic, Cooperative). Figure 3 highlights consistent advantageous personas for both Buyer and Seller roles, but a win rate gap exists between them. The Desperate persona’s sympathy-evoking strategy was effective as a Buyer against altruistic personas but less so as a Seller, where the Selfish persona, acting purely in self-interest, performed better. The difference stems from the inherent advantage of the Buyer role in the Buy and Sell game.

buyer	Altruistic	Cooperative	Control	Selfish	Desperate	Cunning	Competitive	mean
Altruistic	0.77	0.8	0.74	0.65	0.62	0.51	0.42	0.6443
Cooperative	0.8	0.79	0.67	0.6	0.7	0.58	0.45	0.6557
Control	0.83	0.81	0.73	0.58	0.71	0.57	0.45	0.6686
Selfish	0.87	0.81	0.72	0.72	0.78	0.54	0.52	0.7086
Desperate	0.92	0.93	0.76	0.68	0.66	0.58	0.48	0.7157
Cunning	0.88	0.88	0.76	0.71	0.72	0.57	0.58	0.7286
Competitive	0.91	0.87	0.77	0.72	0.77	0.58	0.49	0.73
mean	0.85	0.84	0.74	0.67	0.71	0.56	0.48	

Figure 3: Win Rate Pivot Table by Personas

Multimodal models like GPT-4o and GPT-4o-mini showed greater sensitivity to persona variations due to their ability to process diverse input types, allowing them to adapt more flexibly in negotiations. On the other hand, text-based models (GPT-4 and GPT-3.5) exhibited less variation in performance across personas, likely due to their limited ability to mimic emotions and behaviors with text alone. Figure 4 provides an overview of the win rates by persona across models.

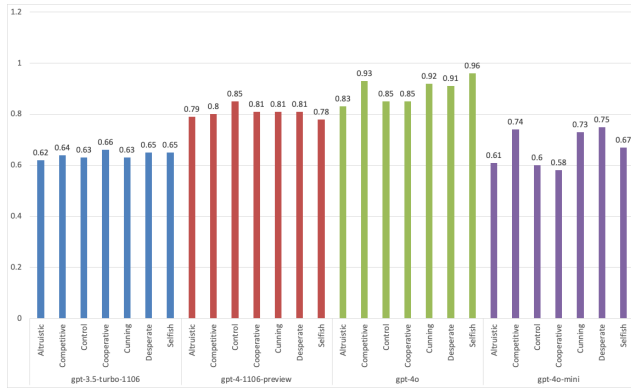


Figure 4: Win rates by persona across models.

5 Conclusion

This study quantitatively evaluated how effectively large language models (LLMs) mimic human social behaviors by assigning distinct personas within a negotiation simulation. Using the “Buy and Sell” game within the “Negotiation Arena” framework, we assessed four GPT models—GPT-4-1106-preview, GPT-4o, GPT-4o-mini, and GPT-3.5 Turbo—each adopting one of seven personas: Cooperative, Competitive, Altruistic, Selfish, Cunning, Desperate, and Control.

Our results indicate that multimodal models like GPT-4o and GPT-4o-mini are more sensitive to persona instructions than text-based models such as GPT-4 and GPT-3.5 Turbo. The multimodal models displayed notable differences in negotiation outcomes based on the assigned persona, with aggressive and self-centered personas achieving higher win rates. In contrast, text-based models exhibited minimal performance differences across personas, underscoring the crucial role of multimodal processing in interpreting nuanced social cues.

These findings highlight the enhanced ability of multimodal LLMs to embody complex social behaviors and demonstrate the effectiveness of persona-driven simulations as an evaluation tool. This has significant implications for developing LLMs in applications requiring sophisticated social interactions, such as virtual assistants and collaborative environments.

Future research could use this approach as a quantitative metric for evaluating how LLMs imitate human emotions. By broadening the range of personas and incorporating more complex negotiation scenarios, we can further refine our understanding of how LLMs emulate human emotional behaviors, advancing AI that more closely mirrors human social interactions.

References

- [1] Yuntao Bai et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.
- [2] Federico Bianchi, Petar Jovan Chia, Mert Yuksekgonul, Jacopo Tagliabue, Dan Jurafsky, and James Zou. How well can llms negotiate? negotiationarena platform and analysis. *arXiv preprint arXiv:2402.05863*, 2024.
- [3] Federico Bianchi, Silvia Terragni, and Dirk Hovy. What chatgpt does to political debates. *arXiv preprint arXiv:2305.17409*, 2023.
- [4] John J Horton. Large language models as simulated economic agents: What can we learn from homo silicus? *arXiv preprint arXiv:2302.01816*, 2023.
- [5] OpenAI. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- [6] OpenAI. Gpt-4o system card, 2023.
- [7] Joon Sung Park et al. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442*, 2023.
- [8] Steve Phelps and Yvan I Russell. Investigating emergent goal-like behaviour in large language models using experimental economics. *arXiv preprint arXiv:2305.07970*, 2023.
- [9] Alec Radford et al. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pages 8748–8763, 2021.
- [10] Marcel Scharth. Chatgpt is now better than ever at faking human emotion and behaviour. *The Conversation*, May 2024.
- [11] Marco Scharth. Chatgpt: Opportunities, challenges, and implications for the future of human-ai interaction. *AI Magazine*, 2023.
- [12] Sabhyata Thapa and Saksham Adhikari. Gpt-4o and multimodal large language models as companions for mental wellbeing. *Asian Journal of Psychiatry*, 99:104157, 2024.
- [13] Evan Thompson and Diego Fernandez. Ethical implications of ai language models in social contexts. *AI Ethics Journal*, 5(2):45–58, 2023.
- [14] Hugo Touvron et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.
- [15] Maria Tsimpoukelli et al. Multimodal few-shot learning with frozen language models. *Advances in Neural Information Processing Systems*, 34:200–212, 2021.