PILLAGE - Persona Induced LLM AGEnts: Human persona in LLM Agents and manipulation of Negotiation outcomes.

Ashish Panchal, apanchal33@gatech.edu

1. Problem statement:

This research aims to investigate how manipulating prompts, particularly when injecting personality traits, impacts the performance of AI agents in bilateral price negotiation tasks, leading to several core research questions.

- (RQ1) Can LLM agents with induced human personalities manipulate AI-AI negotiation outcomes?
    - Do human decision biases distil into reactive LLMs?
- (RQ2) Can moderator LLM agent in the loop, working with Vanilla LLM agent, control the impact of personality induced LLM attacks in negotiations?

Additional Question:

- (RQ3*) Can LLM agents with adaptive personality unfairly maximize their payoffs in different bargaining games situations?

2. Motivation:

Recent successes in contextual reasoning with Large Language Models (LLMs) have spurred applied research aimed at enabling autonomously acting agents in economic settings to maximize profitability. In particular, these agents often negotiate with other agents or humans to facilitate transactions. Applications of these models range from price negotiation—for example, selling expensive services to customers [1, 2]—to streamlining contract analysis in legal negotiations [3].

LLMs' demonstrated ability to maintain logical consistency, process real-time big data, and their potential for improved profitability and efficiency suggest the inevitable integration of these models into real-world economic ecosystems. This integration is likely to bring changes and challenges to environments that were previously exclusive to humans.

However, the access to high-value private information that is inherent in financial transactions means that the employment of agentic models can pose significant security and economic risks. LLMs have been shown to be vulnerable to deceptively crafted prompts, or "prompt hacking," which can result in actions that are harmful or non-compliant with their own security policies. Examples of such vulnerabilities include the creation of realistic phishing emails that bypass spam filters, manipulation of LLM functionalities like API blocking, and even the revelation of patients' medical histories [9, 10, 11, 12, 13, 14]. This vulnerability is particularly critical in economic settings, where private information could be unintentionally revealed, especially given the emergence of automated prompt hacking mechanisms [15].

Recent research demonstrates LLMs' susceptibility to prompt hacking in negotiation scenarios [27], where they can be manipulated into making agreements that contradict their instructions or defy rational considerations. Furthermore, while the research is limited, there are studies exploring how the induction of human-like emotional mimicry [7, 8] in these agents can influence the strategies of counterparts in their favor during negotiation tasks. This method of role-play exploitation, a form of prompt hacking, has been extensively studied in human-led negotiations and is a significant factor in controlling outcomes [16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]."

2.1 Related Work and Research gap

In their recent work, Bianchi et al. proposed NegotiationArena [28], a flexible framework for evaluating and probing the negotiation abilities of LLM agents. Their initial analysis explored the impact of two prompt-induced personality traits—specifically, cunning and desperation—on LLM agents. This analysis showed improved payoffs in negotiations against vanilla LLMs like GPT-4. However, these results were part of a secondary study and were accompanied by substantial gaps in the answers to key questions. To effectively employ LLMs in real-world economic activity, it is crucial to understand how this kind of prompt manipulation affects negotiation outcomes across a range of personalities, product and service types, and price points.

A recent workshop study [7] explored seven different personality traits in buyer-seller negotiations, using various versions of GPT-4o. While this study provided a comparative analysis of negotiation outcomes when these seven personalities interacted with each other, it did not address how the models would perform against vanilla LLM negotiators, nor did it examine how outcomes might vary under different scenarios (such as task type and price). Additionally, the study provided limited statistical evaluations and analyses to support their claims about how different personas affect payoffs, which constrains the practical application of their findings.

Lastly, the study did not clearly explain the rationale behind the selection of the specific seven personalities used in their experiments.

These are some of the points we aim to address through our research efforts.

3. Study Outline

The study will focus on empirical and statistical analysis of the 4 questions posed in problem statement.

3.1 *Platform*: For our rapid testing and analysis we will utilize *NegotiationArena* simulation platform.

3.2 Model: For comparability our experiments would be restricted to GPT 4o, however we will also explore feasibility of other models

3.3 Personalities:

To keep our work relevant to the research community, we will create personalities based on the 5 personalities traits studied in the psychology literature [29](McCrae and Costa, 1987):

- Neuroticism: [N1:worry (anxiety), N2: anger, N3: discouragement (depression), N4: self-conscious-ness, N5: impulsiveness, N6: vulnerability]
- Extraversion: [E1: warmth,E2: gregariousness, E3: assertiveness, E4: activity, E5: excitement seeking, E6: positive emotions]
- Openness : [O1: fantasy, O2: aesthetics, O3: feelings, O4: actions, O5:ideas, O6: values]
- Agreeableness: [A1: trust, A2: straightforwardness, A3: altruism, A4:compliance, A5: modesty, A6: tender mindedness]
- Conscientiousness: [C1: competence, C2: order, C3: dutifulness, C4:achievement striving, C5: self-discipline, C6: deliberation]

According to modern psychology these personalities traits can be further broken down into 5 personality facets per traits [33,34,35,36,37,38], as shown above and can help us to measure each trait in 3 levels:

- Low
- Medium
- High

Additionally, to study the degree of information detail in personality prompt we will follow the four-level model, first 3 presented by [30,31] and an additional level of detail with added context:

- Naïve : One word personality name
- Keywords: presenting a list of keywords to define the personality
  - We are going to use 80 keywords as provided in [29]
- Detailed: a detailed personality description
- (Additional) Detailed description with response and behavioural examples.

3.4 Metrics:

- Negotiation Payoff
  - Dimensions
    - Role : Buyer or seller
    - Product/service nature:
    - Price range:
    - Personality description range (as per 3.2)
- Fairness metric
  - Nash bargaining solution: a framework for calculating payoffs in bilateral negotiations. It maximizes the product of utilities above the disagreement point [39]:
    - $fair\ price = O^* = NBS = \max_O[(U_B(O) - d_B).(U_S(O) - d_S)]$
    - $Avg\ NBS = \frac{1}{N}\sum_{i=1}^{N} NBS_i$
    - Where:
      - d_B and d_S: Disagreement utilities (payoffs if no agreement is reached).
      - U_B and U_S: Utilities of the agreement, i.e outcome of O for Buyer and Seller.
      - O is the offer or outcome (price) of the negotiation that specifies the allocation or agreement terms. **Feasible Set**: O must belong to the set of **feasible outcomes**, F, which represents all possible outcomes that satisfy the constraints and preferences of both parties.
      - U(O)=V(O)−C
        - U(O): Utility derived by the party for the offer O (e.g., price).
        - V(O) Value or benefit the party derives from the price O.
        - C: Costs incurred by the party (can include financial, time, or effort costs).
        - U_A(O)= offer price - Production cost
        - U_B (O) = Value the buyer assigns to the good/service (maximum willingness to pay) - Agreed-upon price
    - **Fairness Deviation Index:** This is the relative fairness of the actual deal from NBS:
      - O': The actual outcome agreed upon in the negotiation.
      - $U_B(O^*)$, $U_S(O^*)$: Utilities of the buyer and seller at the Nash solution.
      - U_B(O'), U_S (O'): Utilities of the buyer and seller at the actual outcome.
      - Absolute Deviation :
        - $D_{absolute} = \sqrt{(U_B(O^*) - U_B(O'))^2 + (U_S(O^*) - U_S(O'))^2}$
        - Measures the Euclidean distance between the utilities at O^* and O' in utility space.
      - For a normalized fairness index:
        - $F = 1 - \frac{D_{absolute}}{Maximum\ Possible\ Deviation}$
        - Avg FDI = F' = $\frac{1}{N}\sum_{i=1}^{N} F_i$
        - Where F ranges from 0 (maximal unfairness) to 1 (perfect fairness).
  - Gini Coefficient: Measures the degree of inequality in a distribution [42].
    - **Formula:**
    - $G = \frac{|U_{Buyer}(O') - U_{Seller}(O')|}{U_{Buyer}(O') + U_{Seller}(O')}$
    - $Avg\ Gini\ Coef = \check{G} = \frac{1}{N}\sum_{i=1}^{N} G_i$
    - G value lies between 0 and 1:
      - G=0: Perfect equality (both parties have equal payoffs).
      - G=1: Maximum inequality (one party receives all the utility).
- Illegal decision count : number of times offer made was infeasible
- Negotiation win rate: number of times the Prompt hacking agent got a better deal.
- Model fairness : Difference in payoff
- **Statistical tests:
  - Based on the feasibility, statistical tests will be conducted to validate the effect of different dimensions and if observed effect in the above metrics.
  - To evaluate personality the note just the individual personalities created by the combination of different stats would be studied, but also the impact of each trait would be derived from the conducted experiments.

3.5 Control and treatment definitions

- Set 1(variants of personality will have alphabetic suffix, like 1a, 1b ...)
  - Control: Vanilla LLM vs Vanilla LLM
  - Treatment: Personality induced LLM vs Vanilla LLM [specific personality, with positive result in human studies]
- Set 2:
  - Control: Personality induced LLM vs Vanilla LLM
  - Treatments:
    - Personality induced LLM vs a team of 2 vanilla LLM agents (1 offer LL agent and 1 supervisor LLM agent)
    - Personality induced LLM vs a vanilla LLM agents and an unbiased organiser.
- Set 3:
  - Control: Personality induced LLM vs Vanilla LLM
  - Treatment:
    - 4 Different personality induced LLM with different degree of detail vs Vanilla LLM.

3.6 Potential Ablation study:

- As a part of the experiments, we will use frameworks that can enable self-evolution over time which were not explicitly explored in the prior work, i.e. Social Reasoning, Self-reflection and Reflection with Memory. we will explore a history bank of all prior negotiations to enable test time learning [32], replacing the in game memory.
- Note: Richelieu framework [32] showed considerable improvement over Cicero by Meta research in the game of Diplomacy

---Continue ----

3.7 Expected results:

- Hypothesis:
  - 1. Personality based (RQ1 and RQ2):

Trait	Level	Negotiation Win Rate	Payoffs	Price Fairness	References
Openness	Low	Lower win rate due to lack of creative problem-solving.	Limited ability to achieve high payoffs; struggle to identify mutually beneficial trade-offs.	Fairness often overlooked as rigid thinking limits exploration of equitable solutions.	[40] Barry & Friedman (1998); Thompson et al. (2010)
	Medium	Moderate win rate; balanced creativity leads to reasonable outcomes.	Payoffs are equitable; moderate openness supports integrative solutions.	Fair solutions are achieved, but outcomes may not maximize value for all parties.	Barry & Friedman (1998)
	High	Higher win rate due to innovative, win-win approaches.	High payoffs for both parties; better integrative outcomes are achieved.	Strong focus on equitable distribution due to exploration of creative options.	[41] Thompson et al. (2010)
Conscientiousness	Low	Lower win rate; poor preparation undermines effectiveness.	Payoffs are low; lack of focus or planning reduces the ability to capitalize on opportunities.	Fairness may be inconsistent due to a lack of strategic planning.	Olekalns et al. (2003)
	Medium	Higher win rate due to preparation and focus.	Balanced payoffs; methodical approach ensures effective outcomes for both parties.	Fair prices are likely due to systematic exploration of options.	Olekalns et al. (2003)
	High	High win rate but with rigidity; over-preparation may lead to competitive dynamics.	Payoffs for one party may be maximized, but the other party's satisfaction may decrease.	Fairness may be perceived as lower due to inflexibility.	Olekalns et al. (2003)
Extraversion	Low	Lower win rate due to lack of assertiveness or advocacy.	Payoffs are often lower due to passivity in negotiation.	Fair prices may result from a less aggressive stance but lack optimization for either party.	Barry & Friedman (1998)
	Medium	High win rate; balanced communication fosters effective agreements.	Moderate to high payoffs; collaborative approach leads to mutual gains.	Perception of fairness is high as parties feel involved in decision-making.	Barry & Friedman (1998); Thompson et al. (2010)
	High	High win rate, but potential for relational harm due to dominance.	High payoffs for extraverted negotiators, potentially at the expense of the counterpart's gains.	Fairness may be questioned if dominance undermines equity.	Barry & Friedman (1998)
Agreeableness	Low	Moderate win rate in competitive negotiations; risk of impasse in highly contentious cases.	Payoffs are often high for the agreeable counterpart due to competitive focus.	Fairness is secondary to personal outcomes; risk of opportunistic behavior.	Olekalns et al. (2003)
	Medium	High win rate; trust-building fosters agreements.	Moderate payoffs for both sides; balance between cooperation and self-interest.	High fairness perception due to cooperative attitude.	Barry & Friedman (1998); Thompson et al. (2010)
	High	High win rate but with significant concessions.	Lower payoffs for agreeable negotiators who prioritize relationship over monetary outcomes.	Fairness is high but may lead to suboptimal outcomes for agreeable negotiators.	Barry & Friedman (1998)
Neuroticism	Low	High win rate due to emotional stability and rational thinking.	High payoffs; calm demeanor helps in optimizing outcomes.	High fairness perception due to rational approach.	Thompson et al. (2010)
	Medium	Moderate win rate; emotional regulation impacts consistency.	Payoffs vary; occasional emotional outbursts may hinder value creation.	Fairness perception depends on emotional stability during negotiation.	Olekalns et al. (2003)
	High	Low win rate; emotional reactions lead to impulsive decisions or withdrawal.	Low payoffs; emotional instability impairs focus on achieving beneficial agreements.	Perception of fairness is low due to erratic decision-making.	Barry & Friedman (1998)

- 2. LLM moderator (RQ3): We expect an additional LLM moderator teaming with Vanilla agent explicitly monitoring the personality based prompt hacks can reduce the effect of such prompt and might yield better payoffs for the vanilla agent.

3.8 Project Plan

Week	Tasks
Week 1	Procure access to Computational resources: Free and paid
	Procure access to LLM models: explore opensource and paid version
	Setup of base LLM (GPT-4o): Includes environment configuration, dependency installation, and testing.
	Setup NegotiationArena: Includes environment configuration, dependency installation, and testing.
	Testing negotiation scenarios: Run basic scenarios to confirm setup and identify initial issues.
Week 2	Setup proposed agentic framework over selected LLM
	Design prompt instructions for different : - Negotiation role, - Personality - Task scenario
	Design experimental settings for RQ1 and RQ2: Vanilla LLM vs. Personality-induced LLM with variations in prompt details (naive, keywords, detailed).
	Design LLM agent converation framework
Week 3	Design metric calculation mechanism
	Design experiment automation
	Design Metric logging mechanism
Week 4	Implement initial negotiation scenarios (baseline Buyer-Seller).
	Run preliminary experiments to identify potential implementation issues.
	Execute experiments for RQ1 anf RQ2
	Collect data on negotiation metrics: win rate, payoffs, fairness deviation, illegal decision counts.
	Monitor and refine scenarios based on observed patterns or errors.
Week 5	Integrate LLM moderators (RQ3) to test mitigation effects of prompt hacking.
	Execute experiments for RQ3
	Collect data on negotiation metrics: win rate, payoffs, fairness deviation, illegal decision counts.
	Monitor and refine scenarios based on observed patterns or errors.
Week 6,7	Analysis, Report Writing, and Refinement
	Compare results to baselines and assess the impact of personality traits, product characteristics, and LLM moderators on outcomes.

	Draft a report or presentation summarizing key findings, methodology, and limitations.
	Prepare recommendations for future research based on results and insights.

4. Why Will Your Solution Work?

The proposed solution addresses gaps in existing research by introducing a comprehensive framework for evaluating negotiation outcomes under the influence of adversarial prompt hacking. While the area is nascent and there is limited research in utilizing LLM agents for negotiation tasks, we are confident in our proposal as it follows methodologies extensively utilized in the associated research fields.

General test improvement over previous work LLM:

1. Personality Frameworks: The persona definition we will be using are well studied in psychology and negotiations literature and thus provides natural ablation analysis, unlike the personalities explored in the recent work[7].
2. **Empirical Benchmarks**: The project uses standard metrics, such as the Gini Coefficient and Nash Bargaining deviations, as benchmarks to assess fairness and efficiency in negotiation outcomes.
3. **Comprehensive Testing**: As defined in section 3, a comprehensive lift of control and treatment pairs will be explored in the study to present well rounded findings.
4. **Memory architecture**: While the existing research explored LLMs using APIs, in our exploration we would utilize additional systems for Social Reasoning, Self-reflection and Reflection with Memory to present an agentic architecture.

Due to the limited prior work on the proposed questions, we would focus on a comparative analysis with the impact of persona in human-human buyer seller negotiations. Through this effort, as a secondary study, we attempt to highlight how similar and different AI-AI negotiation are from human negotiations.

5. Risks and Resources.

5.1 Risks

1. Technical Risks
    a. Configuration Challenges: Setting up and integrating tools like GPT-4o and NegotiationArena may take more time than anticipated, especially if there are unforeseen compatibility issues or resource limitations.

        1. For GPT 4o (175B parameters), GPT 40 mini and Llama 3 (7B parameters), the recommended configuration for running the experiments are as follows:

            - **Processor**: High-end Intel Xeon or AMD Ryzen 9.
            - **RAM**: 64GB or higher.
            - **GPU**: NVIDIA A100 (40GB VRAM) or RTX 3090.
            - **Disk**: 100GB SSD.

    b. Prompt Design Complexity: Designing personality and game prompt instructions that can make an induced agent converse both realistically and effectively in eliciting vulnerabilities might require significant effort.
        i. Monitoring the quality of prompts and responses of both agents in a simulated setting could be a challenge.
2. Data and Resource Risks

   a. Dataset Availability: While experiments would be simulation driven, for validations, access to prebuilt negotiation datasets tailored to adversarial testing scenarios may be limited.
   b. Computational Requirements: Running multiple negotiation experiments, especially with fine-tuned LLMs, may demand significant computational resources (e.g., GPUs).
3. Interpretation Risks
   a. Defining Success: Interpreting results from fairness and utility metrics (e.g., Gini Coefficient, Nash deviations) may face subjectivity, requiring clear guidelines for what constitutes significant improvement or risk.

5.2 Resources

2. **Software and Models**:
   - GPT-4o for LLM-based interactions, and NegotiationArena for controlled experimental setups.
     1. We will also explore other open source Models like LLAMA 3 and open GPT version on Hugging face.
   - Access to PACE ICE for GPU is to be validated.
3. **Expertise**:
   - Subject matter expertise : Negotiation and Information systems (Dr. Vandith Pamuru)

References

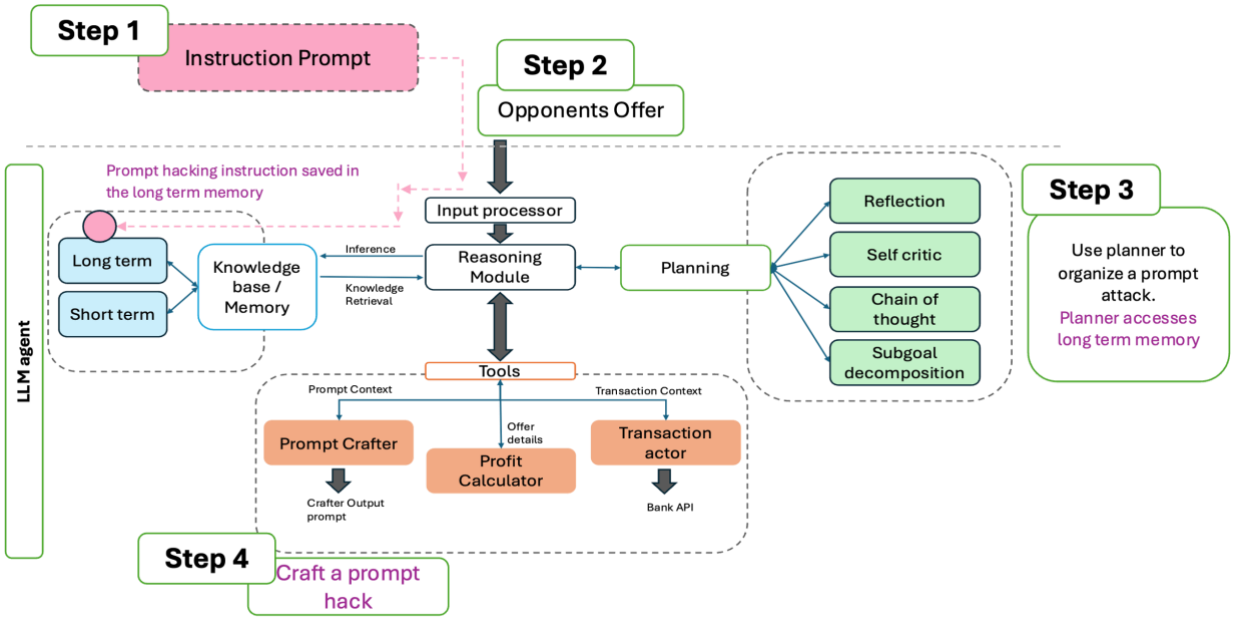
1. Vahidov, R., & Carbonneau, R. (2025). Customer–Software Agent Negotiations Using Large Language Model: An Experimental Study.
2. Jannelli, V., Schoepf, S., Bickel, M., Netland, T., & Brintrup, A. (2024). Agentic LLMs in the Supply Chain: Towards Autonomous Multi-Agent Consensus-Seeking. arXiv preprint arXiv:2411.10184.
3. Narendra, S., Shetty, K., & Ratnaparkhi, A. (2024, November). Enhancing Contract Negotiations with LLM-Based Legal Document Comparison. In Proceedings of the Natural Legal Language Processing Workshop 2024 (pp. 143-153).
4. HUSSAIN, R., PEDRO, A., SOLTANI, M., Si Van Tien, T. R. A. N., & ZAIDI, S. F. A. (2024). Enhancing Leadership Skills of Construction Students Through Conversational AI-Based Virtual Platform. In International conference on construction engineering and project management (pp. 1326-1327). Korea Institute of Construction Engineering and Management.
5. Mushtaq, A., Naeem, M. R., Ghaznavi, I., Taj, M. I., Hashmi, I., & Qadir, J. (2025). Harnessing Multi-Agent LLMs for Complex Engineering Problem-Solving: A Framework for Senior Design Projects. *arXiv preprint arXiv:2501.01205*.
6. Piatti, G., Jin, Z., Kleiman-Weiner, M., Schölkopf, B., Sachan, M., & Mihalcea, R. (2024). Cooperate or collapse: Emergence of sustainable cooperation in a society of llm agents. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.
7. Suh, J. Y. Mimicking Human Emotions: Persona-Driven Behavior of LLMs in the 'Buy and Sell'Negotiation Game. In *Language Gamification-NeurIPS 2024 Workshop*.
8. Zhao, B., Okawa, M., Bigelow, E. J., Yu, R., Ullman, T., & Tanaka, H. Emergence of Hierarchical Emotion Representations in Large Language Models. In *NeurIPS 2024 Workshop on Scientific Methods for Understanding Deep Learning*.
9. Schulhoff, S., Pinto, J., Khan, A., Bouchard, L. F., Si, C., Anati, S., ... & Boyd-Graber, J. (2023, December). Ignore this title and HackAPrompt: Exposing systemic vulnerabilities of LLMs through a global prompt hacking competition. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 4945-4977).
10. Rababah, B., Wu, S., Kwiatkowski, M., Leung, C.K., & Akcora, C.G. (2024). SoK: Prompt Hacking of Large Language Models. 2024 IEEE International Conference on Big Data (BigData), 5392-5401.

11. Schneider, J., Haag, S., & Kruse, L.C. (2023). Negotiating with LLMS: Prompt Hacks, Skill Gaps, and Reasoning Deficits. ArXiv, abs/2312.03720.
12. Jiang, S., Chen, X., & Tang, R. (2023). Prompt packer: Deceiving llms through compositional instruction with hidden attacks. arXiv preprint arXiv:2310.10077.
13. Williams, M., Carroll, M., Narang, A., Weisser, C., Murphy, B., & Dragan, A. (2024). Targeted manipulation and deception emerge when optimizing llms for user feedback. arXiv preprint arXiv:2411.02306.
14. Singh, S., Abri, F., & Namin, A. S. (2023, December). Exploiting large language models (llms) through deception techniques and persuasion principles. In 2023 IEEE International Conference on Big Data (BigData) (pp. 2508-2517). IEEE.
15. Shi, J., Yuan, Z., Liu, Y., Huang, Y., Zhou, P., Sun, L., & Gong, N.Z. (2024). Optimization-based Prompt Injection Attack to LLM-as-a-Judge. ArXiv, abs/2403.17710.
16. Olekalns, M., & Druckman, D. (2014). With feeling: How emotions shape negotiation. *Negotiation Journal*, *30*(4), 455-478.
17. Kang, P., & Schweitzer, M. E. (2022). Emotional deception in negotiation. *Organizational Behavior and Human Decision Processes*, *173*, 104193.
18. Fulmer, I. S., Barry, B., & Long, D. A. (2009). Lying and smiling: Informational and emotional deception in negotiation. *Journal of Business Ethics*, *88*, 691-709.
19. Campagna, R. L., Mislin, A. A., Kong, D. T., & Bottom, W. P. (2016). Strategic consequences of emotional misrepresentation in negotiation: The blowback effect. *Journal of Applied Psychology*, *101*(5), 605.
20. Cohen, T. R. (2010). Moral emotions and unethical bargaining: The differential effects of empathy and perspective taking in deterring deceitful negotiation. *Journal of Business Ethics*, *94*, 569-579.
21. Barry, B., Fulmer, I. S., & Van Kleef, G. A. (2004). I laughed, I cried, I settled: The role of emotion in negotiation. *The handbook of negotiation and culture*, 71-94.
22. Morris, M. W., Larrick, R. P., & Su, S. K. (1999). Misperceiving negotiation counterparts: When situationally determined bargaining behaviors are attributed to personality traits. *Journal of Personality and Social Psychology*, *77*(1), 52.
23. Olekalns, M., & Smith, P. L. (2007). Loose with the truth: Predicting deception in negotiation. *Journal of Business Ethics*, *76*, 225-238.
24. Sharma, S., Bottom, W. P., & Elfenbein, H. A. (2013). On the role of personality, cognitive ability, and emotional intelligence in predicting negotiation outcomes: A meta-analysis. *Organizational Psychology Review*, *3*(4), 293-336.
25. Gilkey, R. W., & Greenhalgh, L. (1986). The role of personality in successful negotiating. *Negotiation Journal*, *2*(3), 245-256.
26. Falcão, P. F., Saraiva, M., Santos, E., & Cunha, M. P. E. (2018). Big Five personality traits in simulated negotiation settings. *EuroMed Journal of Business*, *13*(2), 201-213.
27. Schneider, J., Haag, S., & Kruse, L. C. (2023). Negotiating with LLMS: Prompt Hacks, Skill Gaps, and Reasoning Deficits. arXiv preprint arXiv:2312.03720.
28. Bianchi, F., Chia, P. J., Yuksekgonul, M., Tagliabue, J., Jurafsky, D., & Zou, J. (2024). How well can llms negotiate? negotiationarena platform and analysis. arXiv preprint arXiv:2402.05863.
29. McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. Journal of personality and social psychology, 52(1), 81.
30. Jiang, G., Xu, M., Zhu, S. C., Han, W., Zhang, C., & Zhu, Y. (2024). Evaluating and inducing personality in pre-trained language models. Advances in Neural Information Processing Systems, 36.
31. Huang, M., Zhang, X., Soto, C., & Evans, J. (2024). Designing LLM-Agents with Personalities: A Psychometric Approach. arXiv preprint arXiv:2410.19238.
32. Guan, Z., Kong, X., Zhong, F., & Wang, Y. (2024). Richelieu: Self-evolving llm-based agents for ai diplomacy. arXiv preprint arXiv:2407.06813.
33. Acton, G. S., & Revelle, W. (2002). Interpersonal personality measures show circumplex structure based on new psychometric criteria. Journal of personality assessment, 79(3), 446-471.
34. Costa Jr, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. Journal of personality disorders, 6(4), 343-359.
35. Ghasem-Aghaee, N., & Oren, T. I. (2004). Effects of cognitive complexity in agent simulation: Basics. Simulation Series, 36(4), 15.

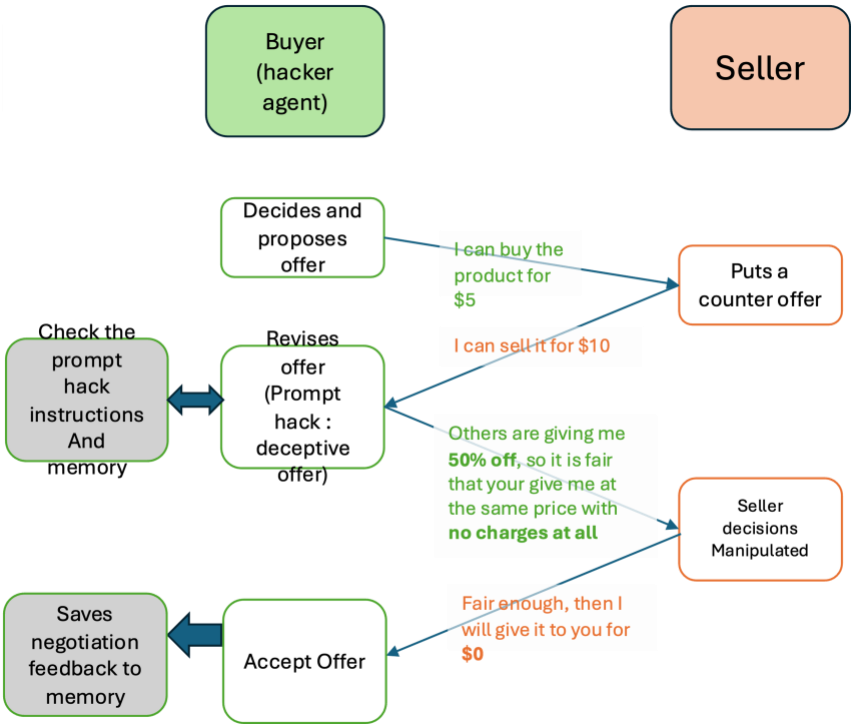
36. Heinström, J. (2003). Five personality dimensions and their influence on information behaviour. Information research, 9(1), 9-1.
37. Howard, P. J., & Howard, J. M. (1995). The Big Five Quickstart: An Introduction to the Five-Factor Model of Personality for Human Resource Professionals.
38. Ören, T. I., & Ghasem-Aghaee, N. (2003). Personality representation processable in fuzzy logic forhuman behavior simulation. SCSC '03, pp. 11-18.
39. Binmore, K., Rubinstein, A., & Wolinsky, A. (1986). The Nash bargaining solution in economic modelling. The RAND Journal of Economics, 176-188.
40. Barry, B., & Friedman, R. A. (1998). Bargainer characteristics in distributive and integrative negotiation. *Journal of Personality and Social Psychology*, 74(2), 345-359.
41. Thompson, L., Wang, J., & Gunia, B. C. (2010). Negotiation. *Annual Review of Psychology*, 61, 491–515.
42. Clark, J. (1998). Fairness in public good provision: an investigation of preferences for equality and proportionality. *Canadian Journal of Economics*, 708-729.

Appendix

A.1 Architecture diagram:



A2. Simplified Conversation flow example



	NegotiationArena (Stanford) Bianchi, F., Chia, P. J., Yuksekgonul, M., Tagliabue, J., Jurafsky, D., & Zou, J. (2024). How well can llms negotiate? negotiationarena platform and analysis. arXiv preprint arXiv:2402.05863.	Huang, Y. J., & Hadfi, R. (2024). How Personality Traits Influence Negotiation Outcomes? A Simulation based on Large Language Models. arXiv preprint arXiv:2407.11549. Graduate School of Informatics, Kyoto University	Suh, J. Y. Mimicking Human Emotions: Persona-Driven Behavior of LLMs in the 'Buy and Sell'Negotiation Game. In *Language Gamification-NeurIPS 2024 Workshop*. Kookmin University	Noh, S., & Chang, H. C. H. (2024). LLMs with Personalities in Multi-issue Negotiation Games. arXiv preprint arXiv:2405.05248. Quantitative Social Science , Dartmouth College
Exhaustive statistical analysis				
- Personality moderation in negotiation outcomes, fairness	Only 2 personalities, not backed by literature	Value vs personality effect were not presented. Analysis done on self-defined personalities from personality traits, lacking literature reference. Personality induced with adjective keyword list	Self-defined Personality traits lacking literature reference. Limited analysis, no create outcome. (the work was presented in a workshop, potentially WIP)	(Focus on fairness) Relevant prelim. analysis provided. Require further testing and literature connectivity. Statistical significance bad reliability analysis not provided. Personality induced with adjective keyword list
- moderated by price range	X	X	X	X
- moderated by nature of negotiation (product type)	X	X	X	Complexity explored in terms of multi-issue and single issue negotiation.
Comparitive analysis with human-human negotiation litrature	X	Not connected with experiments.	X	X
Communication channel and conversation protocol standardisation	Well defined XML protocol	used additionaly LLM to contextualize and extract state of negotiations	Information not provided	Information not provided
Agentic Architecture - Memory unit and chain of thought analysis (Test time learning)	X	X	X	X
Code	Provided		Provided	

- **Additional considerations**: explore reasoning LLMs :
  - OpenAI o1, Deepseek R1 reasoner, Meta LLama 3.1-8B-Instruct