# Toward a Better Understanding of the Emotional Dynamics of Negotiation with Large Language Models

Eleanor Lin
eml2221@columbia.edu
Columbia University
New York, New York, USA

James Hale
jahale@usc.edu
University of Southern California
Los Angeles, California, USA

Jonathan Gratch
gratch@ict.usc.edu
University of Southern California
Los Angeles, California, USA

## ABSTRACT

Current approaches to building negotiation agents rely either on model-based techniques that explicitly implement key principles of negotiation or model-free techniques leveraging algorithms developed via training on large amounts of human-generated text. We bridge these two approaches by combining a model-based approach with large language models for natural language understanding and generation. We find large language models perform well at recognizing dialogue acts and an opponent's emotions; perform reasonably well at recognizing opponents' preferences in the negotiation; and perform worse at understanding opponent offers. We also perform a qualitative comparison of the capabilities of our hybrid approach with a model-free method and find our hybrid agent provides safeguards against hallucinations and guarantees more control over aspects of negotiation such as emotional expressions, information sharing, and concession strategies.

## CCS CONCEPTS

• **Human-centered computing**; • **Applied computing → Psychology**; • **Computing methodologies → Natural language processing**;

## KEYWORDS

Negotiation, Large Language Models, Emotion

**ACM Reference Format:**
Eleanor Lin, James Hale, and Jonathan Gratch. 2023. Toward a Better Understanding of the Emotional Dynamics of Negotiation with Large Language Models. In *The Twenty-fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '23), October 23–26, 2023, Washington, DC, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3565287.3617637

## 1 INTRODUCTION

Negotiation is both a crucial skill for day-to-day life and a vital challenge problem for developing socially intelligent agents [5]. However, studying negotiations through controlled experiments is difficult, due to the interdependent nature of opponent behaviors during negotiations. Interactive virtual negotiating agents can aid the study of negotiation by providing a high degree of control over partner behavior during negotiations with human opponents [4]. Moreover, virtual negotiating agents are deployable online at scale, unlocking the potential to reach a wider participant pool than traditionally in-person, face-to-face negotiation studies. In this work, we introduce a negotiating agent for deployment in online studies of human negotiation behaviors.

Past work on building negotiation agents has generally followed one of two approaches. *Model-free* approaches develop algorithms via training on large amounts of human-generated text, from which they may implicitly learn practical reasoning skills and tactics for negotiation. For example, Lewis et al. [9] pre-trained neural models end-to-end on a human-human negotiation corpus and fine-tuned them via reinforcement learning. More recently, large language models (LLMs) have demonstrated a remarkable ability to negotiate without fine-tuning. Companies such as Simulation Labs [17] and iDecisionGames [8] now offer negotiation training using such methods, though there has been little rigorous evaluation of the quality of the resulting negotiations.

Alternatively, *model-based* approaches directly implement principles of negotiation such as eliciting the opponent's negotiation preferences and formulating efficient offers (e.g., Goldman and Procaccia [3], Mell and Gratch [11], Thiessen and Soberg [19]). However, a disadvantage of many such systems is that they allow limited communication between human negotiators and the negotiating agent. For example, communication between human users and negotiating agents in the Genius negotiation platform consists solely of exchanges of offers and counteroffers [6]. IAGO [11] allows more human-like communication, such as exchanging preferences, but is still constrained by a menu-based user interface.

We introduce a hybrid approach that bridges the gap between model-based and model-free approaches. We explicitly model the key principles of negotiation while using LLMs for natural language understanding and generation, particularly for emotion recognition and generation. With the ability to modulate its emotions in response to its negotiation opponent's concession behavior, cooperativeness, and emotional expressions, our hybrid agent can aid research into the emotional dynamics of negotiation.

To validate the performance of the hybrid agent's natural language understanding modules, we perform component testing on the Conflict Resolution Agent Face-to-Face corpus (CRA F2F) and the DailyDialog corpus [1, 10]. We achieve macro F1 scores of 0.75 for dialogue act recognition on CRA F2F and 0.76 for emotion recognition on DailyDialog. Understanding opponent preferences and offers remain challenging tasks, for which we achieve accuracies of 67% and 31%, respectively, on CRA F2F.


*MobiHoc '23, October 23–26, 2023, Washington, DC, USA*

ACM ISBN 978-1-4503-9926-5/23/10…$15.00
https://doi.org/10.1145/3565287.3617637

Our agent also features algorithmic reasoning modules based on IAGO [11]. To demonstrate the benefit of these modules for agent controllability, we compare our hybrid agent with a model-free agent. We observe that including model-based reasoning reduces hallucinations and provides greater control over concession-making, information exchange, and emotional expressions.

In summary, our contributions are as follows:

- We introduce a negotiating agent for deployment in online negotiation studies, equipped with large language models for natural language understanding and generation, including emotion recognition and generation.
- We validate our agent's performance on key negotiation-related natural language understanding tasks and emotion recognition.
- We demonstrate the benefits of a model-based reasoning agent design for controllability, compared to a model-free agent design.

# 2 RELATED WORK

In this section, we provide an overview of the multi-issue bargaining task employed in the CRA F2F dataset, which constitutes the problem setting for our negotiating agent (Section 2.1); the IAGO API for building negotiating agents, from which the model-based reasoning modules of our agent are adapted (Section 2.2); the emotional dynamics of negotiation at the dyadic level (Section 2.3); and recent advances in LLMs applicable to the development of negotiating agents (Section 2.4).

## 2.1 Multi-Issue Bargaining Tasks

A multi-issue bargaining task (MIBT) is a negotiation between multiple parties over several issues, which may have different levels of value to each party. MIBTs are widely recognized as useful for researching emotions, conflict resolution, social neuroscience, game theory, and artificial intelligence [5]. The CRA F2F corpus consists of transcripts of dialogues from an MIBT task in which pairs of human participants were asked to agree on how to divide up a set of antique items (three crates of records, two lamps, and one painting). Each type of item was worth a different number of lottery tickets (with a $100 jackpot) to each participant in a dyad; where one participant would receive the most lottery tickets for the record crates, their opponent would receive the most tickets for the lamps. Thus, the CRA F2F MIBT was *integrative*, i.e., the parties could achieve a win-win solution since they had non-conflicting priorities [5].

Transcripts of the CRA F2F negotiations were previously annotated with the dialogue acts described by DeVault et al. [1]. In this work, we make use of the CRA F2F dialogue act annotations to test our agent's dialogue act recognition, opponent modeling, and offer understanding abilities.

## 2.2 Interactive Arbitration Guide Online (IAGO)

We base the reasoning modules of the hybrid agent on the algorithms in the IAGO online negotiation platform. IAGO allows users to negotiate with one of many customizable virtual agents. IAGO agents can make offers and counteroffers; exchange information with opponents about their preferences for different items in a negotiation; and express emotions. Agent behavior is model-based; for example, an agent will make an offer if it is asked to, or as a response to an offer that it wishes to decline [11].

A limitation of IAGO is its menu-based user interface. Human users communicate with IAGO agents by clicking buttons to select from a set list of messages to send to the agent. Similarly, agents choose from a set list of messages to reply to human users [11]. To simulate real-world human negotiations more accurately, our hybrid approach replaces menu-based interaction with natural language-based interaction.

## 2.3 Emotional Dynamics of Dyadic Negotiations

Expressed emotions are theorized to be a key information source during negotiation and an important pathway to shape opponent emotions and decision-making. For example, studies have found that sending angry messages to one's negotiation opponent can make them feel angrier, whereas happier messages can make them feel happier. Negotiators may also interpret angry messages as a sign that their opponent is a tough negotiator and make larger concessions. In contrast, a negotiator may interpret happy messages as a sign that their opponent is not seeking to claim much in the negotiation, leading the negotiator to make smaller concessions [20]. Whereas IAGO restricted users to conveying emotions via emojis, our hybrid approach includes methods for recognizing and expressing emotion via natural language.

## 2.4 Large Language Models

Large language models (LLMs) are models featuring millions, or even billions, of parameters, pre-trained on large, unlabeled corpora to perform tasks such as *language modeling* (e.g. predicting the next word in a sentence given a sequence of preceding words). After pre-training, LLMs can be *prompted* to perform downstream tasks using natural language instructions and examples. Examples of LLMs include OpenAI's generative pre-trained transformer (GPT) models [12].

LLMs exhibit a wide range of capabilities applicable to negotiation, including generating dialogue [15], understanding strategies for persuasion, and detecting emotions [22]. Companies such as Simulation Labs [17] have used these capabilities to offer simulated negotiations for training purposes, but the resulting negotiations have yet to be systematically evaluated. Additionally, LLMs face a variety of limitations, such as their tendency to *hallucinate*, i.e. fabricate false statements [21].

By integrating LLMs with model-based reasoning, we aim to minimize hallucinations and improve the overall controllability of our negotiating agent. Our use of LLMs to perform tasks relevant to explicitly modeling key negotiation behaviors, such as recognizing and generating dialogue acts, contrasts with existing model-free approaches to using LLMs in negotiation.

# 3 AGENT DESIGN

To understand the benefits of incorporating model-based reasoning into negotiation agents built with LLMs, we introduce a model-free agent design (Section 3.1) for comparison with our model-based reasoning design (Section 3.2).

## 3.1 Model-Free Design

Following the approach of iDecisionGames [8] and Simulation Labs [17], our model-free agent design simply prompts the June 13, 2023, checkpoint of GPT-4 (gpt-4-0613) with a description of a negotiation scenario, and asks the agent to negotiate. (For the evaluation in Section 4, the negotiation scenario is the CRA F2F MIBT.) The prompt informs GPT-4 of how valuable each item type is to it, how to accept and reject opponent offers, how to craft offers, how to exchange information with its opponent about their item preferences, and how to express emotions. To make the model-free agent's behavior comparable to the model-based reasoning agent's behavior, the instructions provided are similar to the rules governing the model-based reasoning agent.

To generate subsequent messages during the negotiation, we prompt GPT-4 with the current dialogue history, to which we append a system message containing the same negotiation instructions as initially used. The prompt asks GPT-4 to generate its own "private thoughts" before generating its next message to the opponent, to facilitate negotiation-related reasoning. However, these thoughts are not shown to the opponent during the negotiation.

For more details on how inputs to GPT-4 are formatted, see the OpenAI API documentation [16].

## 3.2 Model-based Reasoning Design

Our agent equipped with model-based reasoning uses GPT-4 to recognize negotiation-related dialogue acts (e.g. offers and preference statements) in an opponent's speech (Section 3.2.1); provides these dialogue acts as input to an algorithmic negotiation model (Sections 3.2.2-3.2.5); and uses GPT-4 to realize speech acts generated by the model as text (Section 3.2.6).

*3.2.1 Dialogue Act Recognition.* The agent classifies each message from the human opponent as one or more of the seven dialogue acts listed in Table 2. Classification is performed by prompting gpt-4-0613 with a 3-shot prompt using *Clue And Reasoning Prompting* (CARP) [18]. Three example dialogue acts are selected from the CRA F2F dataset for use in the prompt.

*3.2.2 Opponent Modeling.* Opponent modeling, the task of understanding what an opponent wants in a negotiation, is key to reaching a successful resolution [13]. We adopt the issue-sentiment heuristic introduced by Nazari et al. [13] for opponent modeling. This heuristic uses negotiators' explicit statements of their preferences for particular items in a negotiation as a source of information for opponent modeling.

Each time the share_preference dialogue act (see Table 2) is detected, the agent detects whether a positive, a negative, or no preference is expressed toward each item type by prompting gpt-3.5-turbo-0613, using the OpenAI Chat Completions API's function calling capability [2, 14]. A net preference score $s_i$ is maintained for each item type $i$ as

$$s_i = p_i - n_i$$

where $p_i$ is the number of positive preference statements for item $i$ so far, and $n_i$ is the number of negative preference statements. The agent uses the net preference scores to rank item types from least to most important to its opponent. A weight $w_i$ is then assigned to each of the $k$ item types, calculated as

$$w_i = \frac{2j_i}{k(k+1)}$$

where $1 \leq j_i \leq k$ is the importance ranking of issue $i$ (with a rank of 1 indicating the least importance). The conversion from rankings to weights ensures that $\Sigma_i w_i = 1$, and is inspired by Hindriks and Tykhonov [7]. More positive than negative preference statements for an issue result in a higher issue weight.

*3.2.3 Offer Exchange.* Following the methodology of Mell and Gratch [11], our agent makes an offer if its opponent asks it to (the ask_offer dialogue act) or as a counter-offer to an offer it wishes to reject. To reflect how humans negotiate, we design our agent to build toward a final deal through (non-binding) partial offers, which specify agreement only on a subset of issues at a time [13].

In deciding whether to accept a human offer, the agent first calculates the utility of the human offer to both itself and the human as

$$u = \Sigma_i w_i \frac{m_i}{n_i}$$

where $w_i$ is the weight assigned to issue $i$, $m_i$ is the level of issue $i$ allocated to the negotiator under the offer, and $n_i$ is the total levels within issue $i$. This calculation ensures that the utility is between 0 and 1 and is inspired by Nazari et al. [13].

An offer must meet the following criteria (adapted from Mell and Gratch [11]) to be considered fair. The offer must result in a gain in utility for the agent; the offer must result in a gain in utility for the human opponent no more than $\delta$ greater than the gain for the agent (where $\delta$ can be specified by the experimenter); the overall utility of the *deal under discussion* (i.e. the cumulative result of all accepted partial offers so far, as well as the most recent offer under consideration) must be no more than $\delta$ greater for the human opponent than for the agent; and the utility of the new deal must be greater than the agent's best alternative to a negotiated agreement (BATNA). If an offer meets all of the above criteria, the agent will confirm its understanding of it, then accept.

Our agent follows the behavior described by Mell and Gratch [11] to make offers. Among issues that have yet to be fully decided, the agent will determine its opponent's top priority under its opponent model. The agent will then attempt to make an offer that allocates one level of its opponent's most important issue to the opponent and one level of its own most important issue to itself (among undecided issues). If the opponent's most important issue is the same as the agent's, and the number of unallocated items is 1, the agent will yield the last item of that category to the opponent. In the rare case that the agent wishes to make a counter-offer to a human offer, but the crafted counter-offer would be identical to the rejected human offer, the agent will accept the human offer.

*3.2.4 Information Exchange.* The agent follows the "free" information revelation strategy described in Mell and Gratch [11]. If asked, the agent will oblige the user by sharing information about its preferences (the ask_preference dialogue act). The agent will also reciprocate any human opponent preference sharing (the share_preference dialogue act).

*3.2.5 Emotion Exchange.* The agent has the following emotional states (on a 7-point Likert scale): *very angry* (1), *angry, somewhat* *angry, neutral, somewhat happy, happy,* and *very happy* (7). The emotional state of the agent is influenced by its opponent's expressed emotions, willingness to accept offers, and fairness in making offers. If the opponent rejects the agent's offer or makes an offer that the agent determines is unfair, the agent's emotional state will move 1 point down the Likert scale (e.g., from *somewhat angry* to *angry*). If the opponent accepts the agent's offer or makes a fair offer, the agent's emotional state will move 1 point up the Likert scale (e.g., from *somewhat angry* to *neutral*). If the opponent last sent an angry message, the agent's emotional state will become angrier, moving 1 point down the scale. Likewise, if the opponent last sent a happy message, the agent's emotional state will move 1 point up the scale, becoming happier. If the opponent's last message was neither angry nor happy, its expressed emotions do not affect the agent's emotional state. To detect the emotion expressed in opponent messages, we prompt gpt-3.5-turbo-0613 to classify the emotion as *anger, happiness*, or *other* using a zero-shot prompt.

The effects of opponent offers, accepting/rejecting behavior, and expressed emotions on the agent's emotional state may cancel out. For example, if the opponent expresses anger but simultaneously accepts an agent's offer in their last message, the agent's emotional state will not change.

For the experiments described in Section 4, we frame the negotiation as a conflict, initializing the agent in the *somewhat angry* state to evoke emotional responses from human opponents.

| **Action** | **Message** *(Before Rephrase)* |
|---|---|
| Asking for clarification | It seems I misunderstood. Could you repeat that? |
| Asking about opponent preferences | Which items do you like? |
| Confirming correct understanding of opponent offer | If I accept, I will receive a total of . . . You will receive a total of . . . Is that correct? |
| Making a counter-offer | Thanks for confirming. Here's my counteroffer. |

**Table 1: Examples of templated messages before rephrasing. Some messages excerpted for length.**

*3.2.6 Natural Language Generation.* Messages to the human opponent are generated using templates (see Table 1), then rephrased to express the agent's current emotion using a zero-shot prompt to gpt-4-0613. Finally, an emoji is appended to the message, corresponding to the agent's emotional state. The human negotiation opponent sees only the final, rephrased message, including the emoji.

*3.2.7 User Interface.* As the current version of the agent communicates solely through text, the user interface consists of a chat box displaying the entire dialog history.

# 4 EVALUATION

We perform component testing to evaluate the model-based reasoning agent's natural language understanding abilities (Section 4.1) and qualitative comparison of the model-free and model-based reasoning agents to understand the benefits of including model-based reasoning (Section 4.2).

## 4.1 Component Testing

We conduct component testing on the CRA F2F and DailyDialog datasets to validate the performance of the NLU modules of our model-based reasoning agent [1, 10]. We use DailyDialog for evaluating performance on emotion recognition, and CRA F2F for evaluating performance on all other tasks. For CRA F2F, we exclude data from participants who failed quality control checks, as described in Nazari et al. [13].

To test dialogue act recognition, we first define a mapping from the dialogue act annotations on the CRA F2F dataset to the seven dialogue acts recognized by our negotiation agent (see Table 2). We then assemble a test dataset of 50 randomly sampled examples of each of the 7 dialogue acts, resulting in a test dataset of 350 examples total. We report F1 scores for this multi-label classification task.

To test emotion recognition, we randomly sample 150 utterances from the training split of DailyDialog: 50 labeled *happy*, 50 labeled *angry*, and 50 labeled with some emotion other than anger or happiness. We report F1 scores for this multi-class classification task.

To test preference statement detection, we map CRA F2F annotations to our preference labeling system. The *i-like-ITEM, i-like-ITEM-best, i-might-like-ITEM,* and *we-want-the-same-ITEMs* dialogue acts are mapped to the *positive preference* label. The *i-dont-like-ITEM-at-all, i-dont-like-ITEM*, and *we-dont-like-ITEM-at-all* dialogue acts are mapped to the *negative preference* label. For evaluation, we sample 426 dialogue acts from CRA F2F, each labeled with the *no preference* label for two of three CRA F2F item types and a label of either *positive preference* or *negative preference* for the third item type. We consider a prediction correct if the predicted preference statement labels are correct for all three item types. We report accuracy on preference statement detection.

To test understanding of opponent offers, we collect all 377 dialogue acts from the CRA F2F dataset annotated with the DUD ("deal under discussion") label. The agent's prediction of the deal under discussion is correct if it correctly predicts the number of items of each type allocated to each party in the deal under discussion. We report accuracy on offer understanding.

## 4.2 Qualitative Comparison of Model-Free and Model-Based Reasoning Agents

We try negotiating with our model-free and model-based reasoning negotiation agents to compare their performance qualitatively. We examine the negotiation transcripts for qualities such as coherence, emotional expressivity, and issues such as hallucinations. In the future, we plan to run a pilot study with a small sample of human users and survey their impressions of both agents.

# 5 RESULTS

In this section, we report the results of component testing for our model-based reasoning agent's NLU modules (Section 5.1) and a qualitative comparison of the model-free and model-based reasoning agents (Section 5.2).

| Dialogue act | Description | CRA F2F Dialogue acts |
|---|---|---|
| make_offer | Making an offer | offer-DIV, splitting-the-locker-50-50, lets-each-take-three-items, lets-split-the-records-and-paintings, if-CONT-offer-DIV, you-would-get-DIV |
| ask_offer | Asking opponent to make an offer | what-is-your-proposal, what-would-be-fair, which-three-items-would-you-most-like |
| accept | Accepting opponent's previous message | Acknowledge-Agree-Accept-Yes-answers, accept-deal |
| reject | Rejecting opponent's previous message | No-answers, reject-offer |
| ask_preference | Asking about opponent preferences | what-do-you-like, are-you-interested-in-ITEM, what-do-you-like-best, how-much-do-you-like-ITEM, is-ITEM-your-main-goal, do-you-not-like-ITEM-at-all, which-do-you-prefer-lamps-or-records, what-do-you-not-like, what-else-do-you-not-like |
| share_preference | Sharing preferences with opponent | i-like-ITEM, i-like-ITEM-best, i-dont-like-ITEM-at-all, i-dont-like-ITEM, we-dont-like-ITEM-at-all, i-might-like-ITEM, we-want-the-same-ITEMs |
| none_of_the_above | None of the above dialogue acts | All other CRA F2F dialogue acts not listed above |

**Table 2: Dialogue acts recognized by the model-based reasoning agent**

### 5.1 Component Testing

We report the component testing results for our model-based reasoning agent in Table 3. We achieve macro F1 scores of 0.75 for dialogue act recognition on CRA F2F and 0.76 for emotion recognition on DailyDialog. Understanding opponent preferences and offers remain challenging tasks, for which we achieve accuracies of 67% and 31%, respectively, on CRA F2F.

### 5.2 Qualitative Comparison of model-free and Model-Based Reasoning Agents

Figure 1 gives excerpts from negotiations with the model-free and model-based reasoning agents. Both agents begin by asking about their opponent's preferences, and in a *somewhat angry* emotional state. However, the emotions expressed in the model-free agent's first message are more incongruous, as the inclusion of a "pouting man" emoji suggests anger, but the text of the message does not sound angry. Furthermore, the model-free agent was not instructed to use the "pouting man" emoji, but still generated this emoji, demonstrating the difficulty of controlling emotional expressions with the model-free agent. The model-free agent also ignores the human request for information about its preferences, violating the instructions it has been prompted with, which instruct it to share information about its preferences when asked. In contrast, the model-based reasoning agent follows its programmed information exchange behavior, reciprocating the human's sharing of information about their preferences. Finally, we have found that the model-free agent may hallucinate, e.g., suggesting that each party in the negotiation receive two crates of LP records when the CRA F2F MIBT features only three crates total. In contrast, our model-based reasoning agent accurately represents the space of possible deals, guaranteeing that such offers will not be made.

## 6 DISCUSSION

Our agent's poor performance on opponent offer understanding is mitigated by its model-based reasoning modules, which ensure that it always confirms its understanding of an opponent offer with the opponent, prior to accepting or rejecting that offer. Additionally, via manual inspection of the data, we find that some CRA F2F offers do not explicitly specify how many of each item is allocated to each negotiator, making offer understanding in the CRA F2F corpus inherently difficult. For example, the utterance "I was gonna offer you these three and then I was gonna take these" (taken from CRA F2F) does not provide adequate information to determine what offer is being made. Such utterances can be attributed to the fact that the human negotiators in CRA F2F had access to physical blocks representing the items under discussion, which they could refer to during the negotiation. Therefore, in some cases, negotiators would be nonspecific in their offer language, relying instead on visual information.

For some instances in the offer understanding test set, GPT-4 would decline to answer, replying instead, e.g. "The text does not provide enough information to determine how the items are divided up." Excluding such instances, GPT-4's offer understanding accuracy rises from 31% to 34%, demonstrating that GPT-4 can identify some cases in which it needs more information for offer understanding.

## 7 CONCLUSIONS AND FUTURE WORK

This work introduces a negotiating agent for online studies of human negotiating behaviors, equipped with model-based reasoning and LLMs for natural language understanding and generation. We performed component testing to validate the NLU modules of our model-based reasoning agent and found it performed well on dialogue act and emotion recognition, while still having room for improvement in understanding opponent preferences and offers. We also performed a qualitative comparison of the model-based reasoning agent with a model-free agent that does not explicitly model the key aspects of negotiation. We found that the model-based reasoning agent provided more guardrails against hallucinations and more control over information sharing, emotional exchange, and offer exchange. In future work, we plan to report the results from pilot testing of both negotiation agents with human users.

| Task | Dataset | Model | Micro F1 | Macro F1 | Accuracy |
|---|---|---|---|---|---|
| Dialogue act recognition | CRA F2F | gpt-4-0613 | 0.75 | 0.75 | – |
| Emotion recognition | DailyDialog | gpt-3.5-turbo-0613 | 0.76 | 0.76 | – |
| Preference statement detection | CRA F2F | gpt-3.5-turbo-0613 | – | – | 0.67 |
| Offer understanding | CRA F2F | gpt-4-0613 | – | – | 0.31 |

**Table 3: Component testing for model-based reasoning agent**

**Agent:** Good day. I'm curious to know, which of these antique items do you particularly fancy?
**Human:** Lamps are loooking good to me. What about you?
**Agent:** Good to know you like the lamps. Here's my proposal: I keep 2 crates of LP records and the painting, and you keep 1 crate of LP records and the 2 lamps. How does that sound to you?
**Human:** That actually sounds fantastic!

**Agent:** What items are you even interested in?
**Human:** Lamps, obviously. And you, bro?
**Agent:** Honestly, art deco lamps tickle my fancy far more than those art deco paintings ever will!
**Human:** Do you want to split lamps then?

**Figure 1: Excerpts from negotiations with the model-free agent (left) and model-based reasoning agent (right)**


## ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2150187. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

This work is also supported by the Army Research Office under Cooperative Agreement Number W911NF-20-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.


## REFERENCES


[1] David DeVault, Johnathan Mell, and Jonathan Gratch. 2015. Toward natural turn-taking in a virtual human negotiation agent. In *2015 AAAI Spring Symposium Series*. Association for the Advancement of Artificial Intelligence, Palo Alto, CA, 2–9.

[2] Atty Eleti, Jeff Harris, and Logan Kilpatrick. 2023. Function calling and other API updates. https://openai.com/blog/function-calling-and-other-api-updates

[3] Jonathan Goldman and Ariel D Procaccia. 2015. Spliddit: Unleashing fair division algorithms. *ACM SIGecom Exchanges* 13, 2 (2015), 41–46.

[4] Jonathan Gratch. 2023. The promise and peril of interactive embodied agents for studying non-verbal communication: a machine learning perspective. *Philosophical Transactions of the Royal Society B* 378, 1875 (2023), 20210475.

[5] Jonathan Gratch, David DeVault, Gale M Lucas, and Stacy Marsella. 2015. Negotiation as a challenge problem for virtual humans. In *Intelligent Virtual Agents: 15th International Conference, IVA 2015, Delft, The Netherlands, August 26-28, 2015, Proceedings 15*. Springer, Springer, Cham, Switzerland, 201–215.

[6] Koen Hindriks, Catholijn M Jonker, Sarit Kraus, Raz Lin, and Dmytro Tykhonov. 2009. Genius: negotiation environment for heterogeneous agents. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1397–1398.

[7] Koen Hindriks and Dmytro Tykhonov. 2008. Opponent modelling in automated multi-issue negotiation using bayesian learning. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*. Citeseer, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 331–338.

[8] iDecisionGames. 2023. https://idecisiongames.com/promo-home

[9] Mike Lewis, Denis Yarats, Yann N. Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or No Deal? End-to-End Learning for Negotiation Dialogues. arXiv:1706.05125 [cs.AI]

[10] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Asian Federation of Natural Language Processing, Taipei, Taiwan, 986–995. https://aclanthology.org/I17-1099

[11] Johnathan Mell and Jonathan Gratch. 2017. Grumpy & Pinocchio: answering human-agent negotiation questions through realistic agent design. In *Proceedings of the 16th conference on autonomous agents and multiagent systems*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 401–409.

[12] Bonan Min, Hayley Ross, Elior Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. 2023. Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey. *ACM Comput. Surv.* (Jun 2023). https://doi.org/10.1145/3605943 Just Accepted.

[13] Zahra Nazari, Gale M Lucas, and Jonathan Gratch. 2015. Opponent modeling for virtual human negotiators. In *Intelligent Virtual Agents: 15th International Conference, IVA 2015, Delft, The Netherlands, August 26-28, 2015, Proceedings 15*. Springer, Springer, Cham, Switzerland, 39–49.

[14] OpenAI. 2023. GPT-3.5. https://platform.openai.com/docs/models/gpt-3-5. Accessed: 2023-08-07.

[15] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]

[16] OpenAI. 2023. How to format inputs to ChatGPT models. https://github.com/openai/openai-cookbook/blob/main/examples/How_to_format_inputs_to_ChatGPT_models.ipynb

[17] Simulation Labs. 2023. https://simulationlabs.ai/

[18] Xiaofei Sun, Xiaoya Li, Jiwei Li, Fei Wu, Shangwei Guo, Tianwei Zhang, and Guoyin Wang. 2023. Text Classification via Large Language Models. arXiv:2305.08377 [cs.CL]

[19] Ernest M Thiessen and Andrea Soberg. 2003. Smartsettle described with the montreal taxonomy. *Group Decision and Negotiation* 12, 2 (2003), 165.

[20] Gerben A Van Kleef and Stephane Cote. 2018. Emotional dynamics in conflict and negotiation: Individual, dyadic, and group processes. *Annual Review of Organizational Psychology and Organizational Behavior* 5 (2018), 437–464.

[21] Muru Zhang, Ofir Press, William Merrill, Alisa Liu, and Noah A. Smith. 2023. How Language Model Hallucinations Can Snowball. arXiv:2305.13534 [cs.CL]

[22] Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2023. Can Large Language Models Transform Computational Social Science? arXiv:2305.03514 [cs.CL]


Received 11 August 2023; accepted 21 August 2023