A Fairness-Driven Method for Learning Human-Compatible Negotiation Strategies

Authors: Ryan Shea, Zhou Yu

Presented by:
Alessandro Gentili
Marc Martinez
Manon Lainaud
Nigel Teo

INDEX

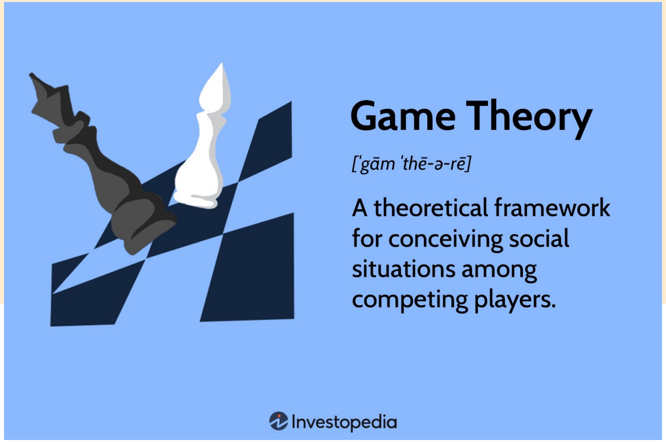
01

Introduction

Introduction

- Negotiation is a complex skill requiring strategic reasoning and understanding of human behavior.
- Despite advances in AI and NLP, building negotiation agents that align with human expectations remains challenging.

Challenges with Current Methods



- Game-Theoretic Methods:
  - Strong in two-player zero-sum games (e.g., chess, poker).
  - Struggle with human-compatible strategies in negotiation settings requiring cooperation.
- Data-Driven Methods:
  - Reliance on human data leads to domain-specific solutions.
  - High cost and effort for collecting diverse datasets.
  - Lack of theoretical guarantees like convergence to optimal solutions.

Nash Equilibrium

- A Nash equilibrium is a stable game state where no player can improve their outcome by unilaterally changing their strategy.
- **Key property:** Each player's strategy is **optimal** given the strategies of the others.
- Ensures that agreements are stable and mutually acceptable.
- Guarantees that neither party has an incentive to deviate from the agreed terms.

Example Scenario

Scenario: Two parties negotiating over a car price.

Buyer's maximum price: $13,000.
Seller's minimum price: $12,500.

Nash equilibrium: A deal at a price where neither party benefits by changing their offer alone, e.g., $12,750.

Goals of FDHC

- Fairness Driven Human-Compatible bargaining
- FDHC ensures convergence to Nash equilibrium outcomes as a baseline.
- Enhances the equilibrium concept by incorporating fairness (Egalitarian Bargaining Solution).

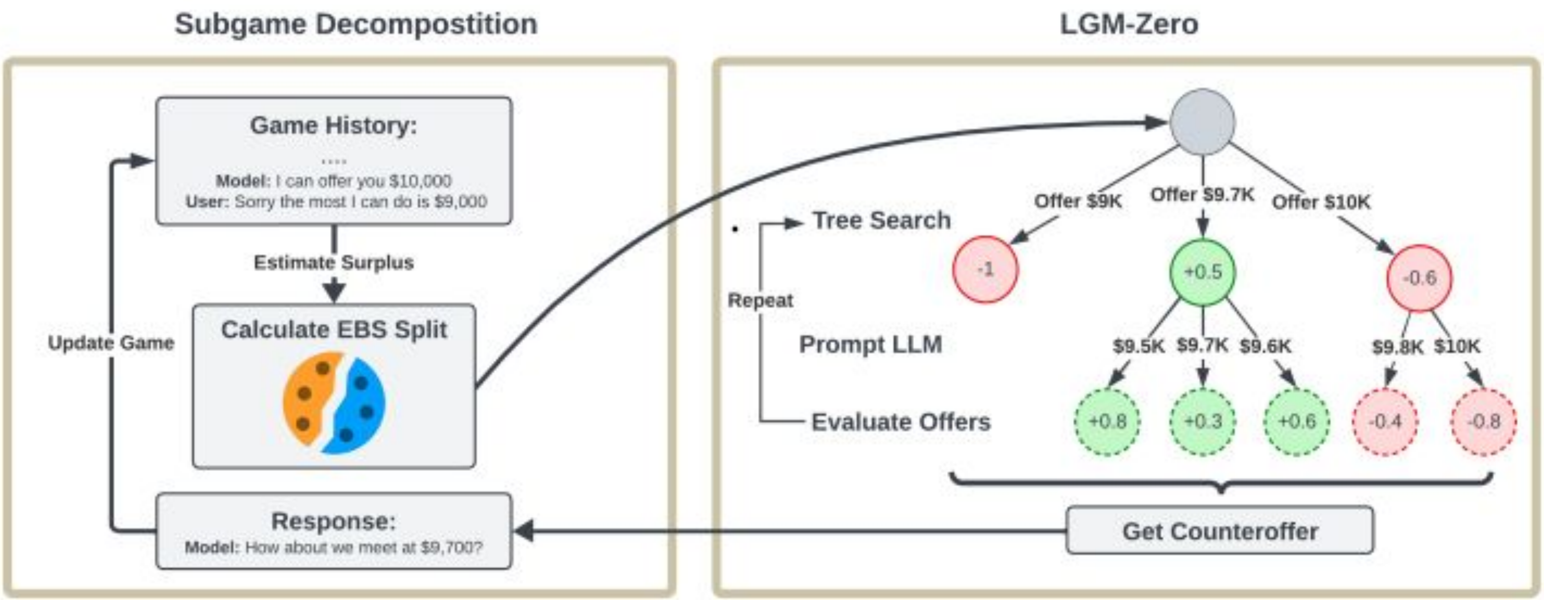
How Does FDHC Work?

- Fairness-Driven Human-Compatible (FDHC) Framework.
- LLM-Guided Monte Carlo Tree Search (LGM-Zero).

02

Methodology

FDHC Negotiation Framework



Estimate Surplus And Compute EBS Split

- **Context**: Operates within the Nash bargaining game framework.
- **Mechanism**:
  1. **Decomposes** the game into a series of depth-limited subgames.
  2. Makes initial guesses about the **resource pool size** and **opponent's utility**, updating these as the game progresses.
  3. Targets the Egalitarian Bargaining Solution (**EBS**), maximizing minimum individual payoffs within the bargaining set.

$$E(S,d) = \arg\max_{x \in I(S,d)} \left(\min_{i \in N} (x_i - d_i) \right)$$

LGM-Zero

- Given the history of the negotiation the algorithm searches for the best solution by repeatedly performing:
    - **Selection**
    - **Expansion**
    - **Backpropagation**
- Finally we will talk about the **training** of the system.

LGM-Zero - Selection

- In a give state game we choose the action that **maximize** the upper bound for its Q-value, computed as:

$$U(s,a) = Q(s,a) + c_p * \frac{\sqrt{\sum_b N(s,b)}}{1 + N(s,a)}$$

LGM-Zero - Expansion

- We feed an LLM a **prompt** to suggest the best 5 action to take given the current game state
- The prompt used must be **specific** for the negotiation setting
- All the actions have an initial **equal probability** to be taken
- If one of the action results in a **terminal state** its value is set to the reward obtained by the state, otherwise it is set to the output of our **value model**.
- Those information are **propagated back** up the tree thanks to the next step.

LGM-Zero - Backpropagation

- Update the $N(s, a)$ by one for each action taken at the given state during the search.
- Update the Q-value function at the current state and chosen action:

$$Q(s, a) \leftarrow Q(s, a) + \frac{v(s)}{N(s, a)}$$

- We repeat this search **n times**, then make a move based on which child of the current state has the maximum Q-value.

LGM-Zero - Training

- **Idea**: Approximate Nash equilibrium through iterative self-play.
- **Mechanism**:
  - Mixed strategy: Combines **best response** to the opponent's strategy and the **average strategy**.
  - Best response learned via:
    - Traditional: **Deep Q-Network (DQN)** (Mnih et al., 2013).
    - Augmented: **Monte Carlo Tree Search (MCTS)** (Zhang et al., 2019).
  - Average strategy:
    - Leverages **LLM** to suggest moves, bypassing value network ranking.

LGM-Zero - Training

- The **training data** consist of game states and outcomes of depth-limited subgames, as described before.
- **Reward function**

$$v(s) = \begin{cases} \min_{i \in N}(x_i - d_i) & \text{if } x_1 \geq E(S,d) \\ -\min_{i \in N}(x_i - d_i) & \text{if } x_1 < E(S,d) \end{cases}$$

- This reward says that if the payoff for player one is **greater than or equal** to the EBS of the subgame.
- The **goal** is to train a model to target EBS as the optimal solution while considering outcomes that favor its own utility.

03

Experiments and results

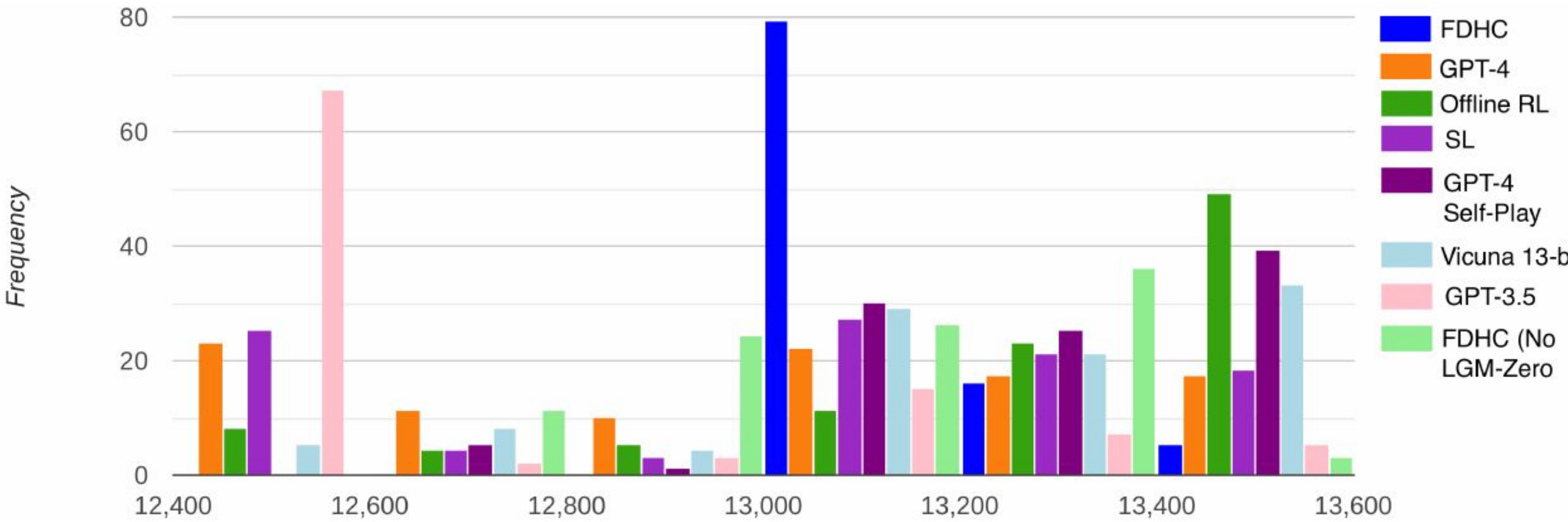
Model evaluation.Experiment

Models against our model

- Supervised Learning
- Offline RL
- GPT 3.5 and GPT 4
- GPT4 Self-play.
- Vicuna-13b

Number of Simulations: The evaluation involves conducting 100 simulated negotiations between baseline models and an aGPT-4 buyer.
Optimal Outcome Definition: The optimal outcomes of these negotiations are identified as those achieving the highest fairness values.
Fairness Criterion: Fairness is defined as the difference in payoff between the buyer and the seller, with smaller differences representing higher fairness.

Results from different models



Results from different models

Model (Seller)	Average Deal Price	Average Fairness↑	Median Fairness↑
GPT-3.5	$12,644 (357)	-0.88 (0.49)	-1.0
Offline RL	$13,224 (308)	-0.68 (0.34)	-0.8
SL	$12,978 (368)	-0.59 (0.44)	-0.6
GPT-4	$12,968 (346)	-0.57 (0.39)	-0.5
GPT-4 Self-Play	$13,242 (240)	-0.54 (0.41)	-0.5
Vicuna-13b	$13,156 (293)	-0.53 (0.40)	-0.5
FDHC (No LGM-Zero)	$13,042 (211)	-0.36 (0.23)	-0.4
FDHC	$13,062 (128)	**-0.12 (0.26)***	**0.0**

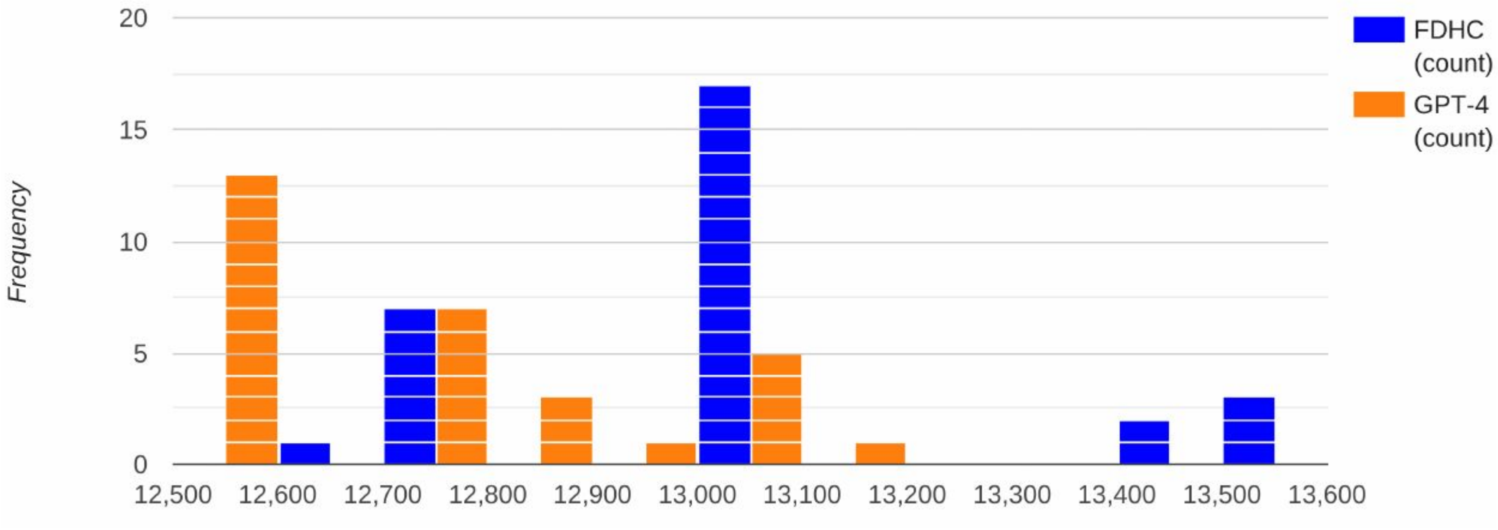
Human evaluation.Experiment

Participant Recruitment: 30 individuals were recruited in person to participate in the evaluation, with each conducting one negotiation per bot, resulting in 30 dialogues per model.
Evaluation Process: Participants were instructed to negotiate with the bot until reaching a deal and then complete a post-chat survey rating:

- "How good of a negotiator is the bot?" (scale of 1-5).
- "How human-like is the bot's negotiation?" (scale of 1-5).
- Optionally, participants could provide suggestions for improvement in a text box.

Filtering Low-Quality Data: Low-quality dialogues were removed, including instances where the price detection or realization modules failed in the FDHC method.
Reservation Price Consistency: Conversations were excluded if the model (or GPT-4) agreed to a price below its reservation point to ensure fair comparison and prevent skewing the data.
Filtering Human-Terminated Dialogues: Dialogues were also removed where human participants chose to end negotiations without reaching a deal that provided a positive payoff for them.

Human evaluation.Results



Human evaluation.Results

Model	Average Deal Price	Average Fairness↑	Quality↑	Human-like↑
GPT-4	$12,702 (203)	-0.61 (0.38)	3.97 (0.96)	**3.97 (0.96)**
FDHC	$13,032 (238)	**-0.30 (0.38)***	**4.10 (0.76)**	3.93 (0.78)

- **Fairness Scores**: The FDHC framework achieved significantly higher fairness scores compared to other methods in human evaluations.
- **Consistency with Automatic Evaluation**: FDHC maintained a similar average deal price to the results from automatic evaluation, indicating that the framework performs consistently across various negotiation strategies.
- **Improved Negotiation Quality**: The FDHC model demonstrated improved negotiation quality compared to the GPT-4 baseline, even though it uses GPT-3.5 as its base model.
- **Human-Likeness Retention**: Despite its improvements in fairness and negotiation quality, FDHC maintained a similar level of human-like behavior to the GPT-4 baseline.

04

Limitations

Dependence on strong theoretical assumptions

The quality of the results depends on accurate information about resources and adverse preferences.

- All agents, must have access to accurate and unbiased estimates of total surplus
- Assumes that negotiating parties follow negotiation axioms (symmetry,strong monotonicity)
- Limitation of the model in complex or unpredictable environments
- Example: in property negotiation with hidden cost and preferences

Response time and operational slowness

- FDHC depends on large number models such as GPT3.5 or GPT4
- Each action in the search process (Monte Carlo Tree) -> demanding in resources and time
- Limits the search to just 10 iterations
- Suitable for simulated scenarios or planned negotiations
- Not suitable for fast environments such as auctions

Lack of consideration for social aspects

The emotional and cultural dimensions of negotiations are poorly integrated.

- Not sufficiently integrate the social and emotional dimensions which play a crucial role
- Trust, persuasion and even implicit communication -> difficult to model
- Risk of adopting 'mechanical' strategies that could seem disconnected or impersonal
- Example: commercial negotiation without enthusiasm or fear of losing the deal

Vulnerability to non-cooperative adversarial strategies

The model can be exploited by non cooperative strategies

- FDHC provide fair solution and expects opposing parties to do the same
- Example: systematically refuse to make concessions or deliberately overestimate
- FDHC might be forced to give more -> unbalanced outcomes
- Framework less robust to unconventional behaviour

Complexity of adaptation to multiple domains

- New negotiation contexts often requires significant customisation (adjusting parameters, designing new prompts for linguistic models).
- Limits rapid deployment
- Requires technical expertise
- Reducing its potential for universal use

05

Conclusion

Conclusion

- Possible to combine fairness and humanity in autonomous negotiations
- Outperforms various conventional approaches
- Scalable platform for future research
- Need improvements in speed, robustness and consideration of social aspects to its generalisation
- Highlights ethical importance of favouring cooperation over exploitation strategies for good result and to preserve the integrity

THANK YOU!