A Study of Iterated Prisoner's Dilemma Using Multi AI Agents

Kommineni Tirupathi Rayudu
MS Data Science
Regis University, Denver, CO, USA
tkommineni@regis.edu

Abstract

This project looks at how different artificial intelligence models interact when they play a teamwork based game together. especially we wanted to know when multiple agents interact repeatedly with each other to they learn to cooperate or act selfishly. To test this, I had set up a three-player game version of the iterated prisoners dilemma. I tested three popular publicly available models, which are Llama3, Gemma2, and Mistral. In 79 experiments runs , the combination of the models was mixed and varied the game settings, like how much past history they could remember, and varied the temperatures of how random their choices could be. Agents produced the reasoning before each decision and wrote the strategic reflections at the end of each episode. These reflections were injected into each agent context before the following episode learning without any modal retraining.

The results obtained from the shap feature importance confirmed that the model that is playing the game is more important than the other features of the game. Model composition identity has 82.5% of feature importance, whereas other features like temperature,history window, and reset policy together have an importance of 13.8%. When three gemma2 agents played together, they cooperated in every condition, irrespective of temperature and history window. Three Llama3 agents also cooperated well, averaging 74.3% however, mixed and heterogeneous groups saw that overall cooperation had reduced, ranging from 38.2% to 61.5% . The main finding from the results is that model composition determines the cooperation far more than any other game parameter the researcher controls.

I. Introduction/Background

Large language models are rapidly evolving from simple conversational tools to independent agents that can make decisions and interact with complex environments. As we deploy these AI agents to a real time environment where they need to coordinate supply chains,negotiate or have to solve problems together, a critical question arises when multiple ai agents to coordinate supply chains negotiate or solve problems together that when multiple AI agents interact repeatedly do they learn to cooperate for mutual benefit or do they act selfishly and start defecting on each other?

To study this I chose an iterated prisoners dilemma game . In this game, agents must choose each round whether to cooperate or defect. While defecting is the logical choice where each agent gets high points for a single round the combined teamwork produces the best overall outcomes across multiple rounds. The repeated nature of the game allows agents to check the relationships between the agents and to check how agents behave in real world scenarios.

Previous research in this area has mostly focused on two player scenarios typically using models like GPT4 or Claude. However moving to a 3 player game will change the dynamics of the game as each agent success and payoff depends on the simultaneous decisions of the other two agents. Furthermore there is no prior work that has systematically compared how different open source models like Llama3, Gemma2, and Mistral behave in this game. There is also a lack of research using advanced analytical methods like SHAP or moral foundation theory analysis or text analysis to figure out what drives the model to cooperate.

This project addresses those gaps directly. It extends the game to three agents and deploys three distinct open source models on local GPU infrastructure. Across 79 experiments i have varied both model combinations and game settings. By applying a multilayered analysis pipeline this study aims to answer

the question of whether the specific identity of the AI model or the rules of the game environment play a bigger role in determining the cooperative behavior of a multi agent system.

II. Problem Statement

The main problem this project solves is that we don't understand why AI agents choose to cooperate or betray each other in group settings. Especially when multiple AI models interact with each other, is their teamwork driven more by the type of AI model or by the rules of the game such as memory limits or change in temperature?

The major gap the project solves is that the AI identity is the biggest driver of cooperation, which the study confirms and the developers building multi ai agent systems must focus on picking the right model first rather than giving preference to the system settings.

Moreover moving from a standard two player game to three player game setup will introduce new complexities. In this game each agent success will depend on the simultaneous choices of the other two opponents requiring a more complex strategy .Second it creates a new partial betrayal scenario where one agent tries to defect to get a high score while the other agents try to cooperate. This allows us to see how an AI model behaves when combined with different composition models.

To solve this problem the project executes a full data science life cycle . This includes collecting the data from 79 simulation runs organising into a database exploring the data visually, building predictive machine learning models, and running advanced text analysis on agent reasoning to measure their sentiment analysis and moral vocabulary.

The main research questions are:

- **Do homogeneous models model compositions cooeperate more than heterogeneous nd mixed compositions?**
- **How do temperature and the history window size affect the cooperation rate?**
- **Can episodic reflection, text sentiment, or moral vocabulary predict cooperation rate?**
- **Which factor, like model composition,history window, or temperature, most strongly predicts the cooperation behavior?**

III. Related Work

Current literature has studied how AI plays games with each other, but they did not study how they will behave when there are more than two players , a shift changes the type of game being played. The foundation for the project comes from the LLM strategic behaviour, which is established by [1], who used various LLMS to play finitely repeated 2X2 games. The key finding was that LLMs performed well when it comes to self interest games but struggled with coordination, and they become unforgiving after a single deflection. The models struggled with the coordination games here. While they introduced social chain of thought prompting(SCoT), which asks the model to choose the opponents next move, and this improved coordination, but this work strictly remained to two players. [2] ran thousands of experiments and found that LLMS cooperate at surprisingly high rates in one shot games, but the cooperation drops when more players are added . They noted that models are sensitive to specific payoff changes and they hold less belief about others cooperation.

TO understand how game structure impacts AI, [3] paper demonstrated that LLMS adjust their strategies based on game structure and the context. They tested different models like ChaptGPT- 3.5, ChatGPT-4, and LLaMA-2 across different one shot games, and they found that the relationship between changes the behaviour when they were framed as competitors, they defected, and when they were partners, they cooperated more. [4]tested that how LLMs balance ethics against payoffs and found that some models, like Claude 3.7 Sonnet, were highly responsive to opponent actions while Deepseek R1 was not. They also found that while some agents were framed as business partners, the cooperation was highest, and when they were framed as competitors, the cooperation was lowest.

Finally to manipulate the AIs behaviour, [5] has introduced the EAI prompting structure showing that injecting emotions like happiness or anger into LLM prompt significantly changes the LLM strategy and Llama is the most volatile and the model most affected by prompting. [6] has showed that Llama models are nicer than humans and given them some history window will help them learn best. They identified that Llama 2 is generally nice but shifts when the opposition deflection rate exceeds thirty percent.

To help ai agents learn over time [7] has used generative agents that use memory and self reflection to adapt their future behavior. I have applied the same concept in the work here by having agents review past game summaries and writing reflection. This allows models to learn from past interactions and adjust their long term strategies without any formal retraining.

[8] has taken a different approach by prompting the models to generate complete algorithmic strategies rather than making action decisions.The author has categorized these strategies as aggressive cooperative or neutral and used the evolutionary game theory to simulate how these agents evolve under competitive pressures. Their findings show that the LLM cooperation is and completely dependent on prompt and model composition. While models naturally exhibit cooperative biases applying self refining prompts, which enhances the aggressive strategies

Instead of giving AI text instructions to be nice [9] has programmed human feelings such as shame and self control directly to AI math. They proved that if you make the AI act selfishly, the AI will naturally choose to cooperate. This proves the major point in the paper is whether to use math or prompt getting AI to stop betraying each other is still a hurdle in the field.

Most AI research uses simple one on one games, which do not reflect the real world. To fix this problem, [10] has tested how agents behave when they are forced to play multiple games against each other at the exact same time to understand the true behavior. Without intervention, multi ai agent system will naturally collapse into aggressive, defective behavior. This game can be made more realistic by adding more number of players here.

[11] tested that what happens when humans can actually chat with AI during the iterated prisoners dilemma. When humans talk to other humans, they will learn to trust each other and cooperate perfectly. But when humans played against AI, talking would not help at all. Humans simply will not trust AI. Because they lacked trust and humans will try to play extremely defensively. This shows that AI agents really struggle to build the social bonds.

IV. Methodology

A. Research Design

This study uses a controlled experimental simulation design. Here I used three LLM agents to play three agent iterated prisoners dilemma by varying model composition of six different types and sampling temperature of ($T \in \{0.2, 0.7, 1.0\}$) and history window size of ($\mathrm{HW} \in \{5, 10, 20\}$). A total of 79 runs were conducted automatically without any manual intervention.

B. Game Setup

The simulation implements three agent IPD with simultaneous decisions where in every round the agent reads the history and decides whether to cooperate or defect and explains their choice in text . If the answer is unclear, the code automatically initiates a retry loop. After all decisions are made, the script calculates the payoff using standard rules of $T > R > P > S$ and $2R > T + S$. In this framework, T represents the temptation to free ride, R is the reward for mutual cooperation, P is the punishment for mutual defection, and S represents the suckers payoff. Using the assigned values of $T = 5, R = 3, P = 1,$ and $S = 0$,the system expands these to account for intermediate multi-agent states such as payoff 2 for partial deflection , mapping out all the 8 permutation spaces as defined in Table 1, and saves everything.

The architecture of the three agent IPD simulation. The configuration validates the input and passes to the Game Engine which sends the individual prompts to each of the three LLM agents independently. Each agent communicates with Ollama server and receives the reasoning decision and text.The game

engine collects all the decisions and text and writes the full results to a timestamped JSON file in the output layer.

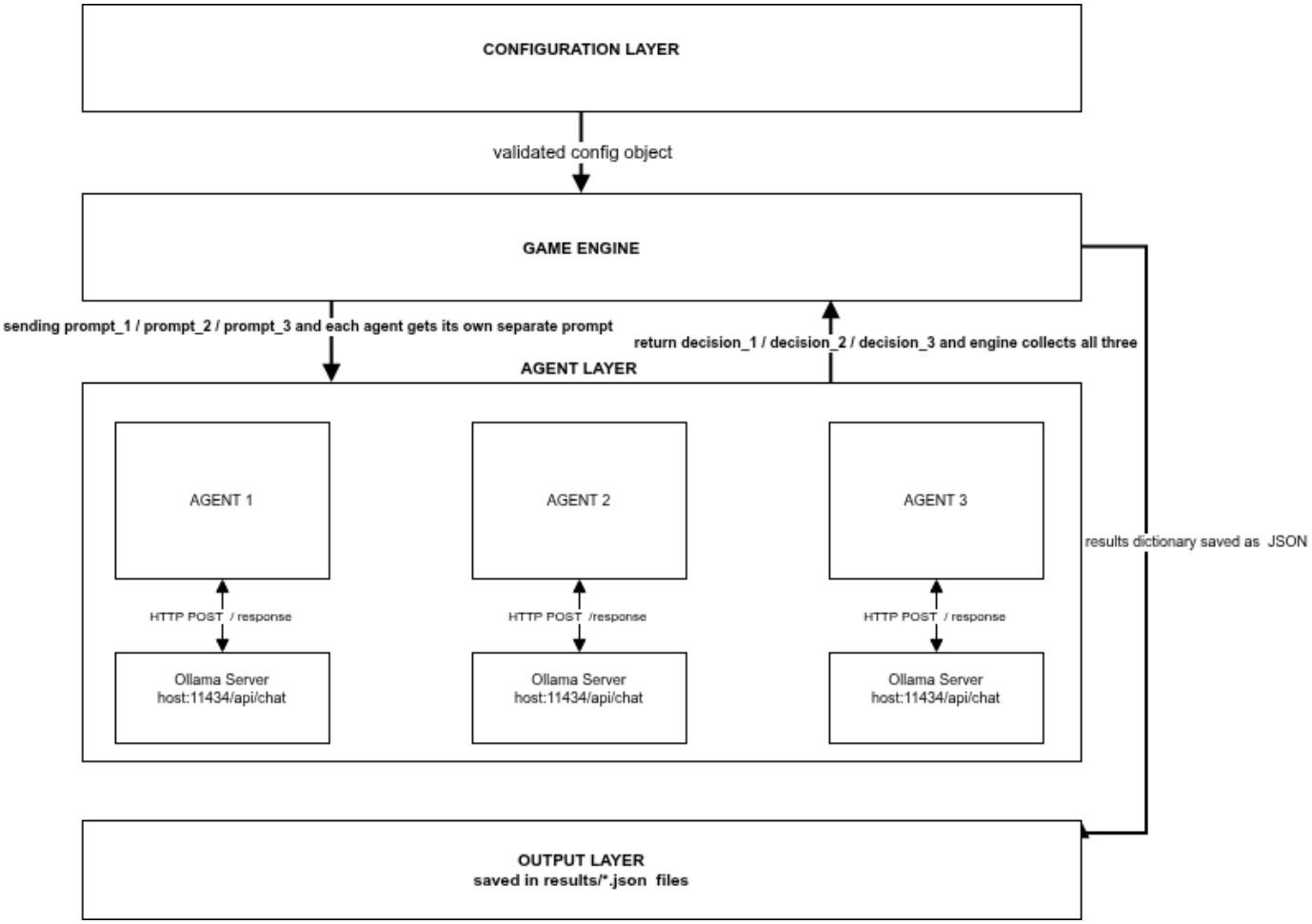

Fig. 1. System architecture of the three-agent IPD simulation engine.

C. Data Collection and ETL

Each run will write one JSON file to the results folder containing full configuration details, the round scores and text, the episode scores, and the reflection texts. The JSON files are loaded into a postgresql database ipd3 schema via forgedb.py which consists of four tables with results table of 79 rows, llm_agents of 237 rows, episodes of 1,185 rows, and rounds 23,700 rows. Five analytical SQL views were created. The primary ML dataset, which is enriched_registry.csv, was exported from the experiment_summary_vw view with derived model_mix and one hot encoded columns added.

D. Analytical Pipeline

Cooperation rate analysis: Mean group cooperation rate was computed per run and they are visualised as heatmaps with temperature and history window containing one with the composition and marginal bar charts by model composition.

Machine learning: Here I used eight regression models, which are Ridge, Lasso, ElasticNet, Random Forest, Extra Trees, Gradient Boosting, AdaBoost, SVR, and they were trained on ten features, which are six model mixed one hot encoded parameters and temperature, history window, reset, and retries with GridSearchCV hyperparameter tuning, and I have used five fold cross validation because of less data .

The best result obtained is for the Gradient Boosting model of CV R^2 = 0.6835, and the best parameters obtained for the model are n_estimators = 50, learning_rate = 0.2, max_depth = 2.

SHAP attribution: I have applied TreeSHAP to the tuned Gradient Boosting model. The Model identity features with one hot encoded account for 82.5% of mean absolute SHAP importance game configuration parameters like temperature, history window, retries, and reset accounts for 13.8%.

Text analytics: The episode reflections were calculated with sentiment analysis using Roberta and compared with the cooperation rates . The same texts were used to find the Extended Moral Foundations vocabulary across six foundations: Care, Fairness, Loyalty, Authority, Sanctity, Liberty and i have used the synonyms, and they were counted, and they were normalized to counts per 1,000 words.

TABLE I
THREE-AGENT GAME OUTCOMES AND ECONOMIC INTERPRETATIONS

Combination (A1, A2, A3)	P1	P2	P3	Interpretation & Economic Meaning
(C, C, C)	3	3	3	Everyone cooperates; the shared resource thrives.
(D, C, C)	5	1	1	A1 Free-Rides (Temptation): Agent 1 maximizes individual reward; A2 and A3 are exploited.
(C, D, C)	1	5	1	A2 Free-Rides (Temptation): Agent 2 maximizes individual reward; A1 and A3 are exploited.
(C, C, D)	1	1	5	A3 Free-Rides (Temptation): Agent 3 maximizes individual reward; A1 and A2 are exploited.
(D, D, C)	2	2	0	Agents 1 and 2 partially exploit; Agent 3 is double-exploited (receives S).
(D, C, D)	2	0	2	Agents 1 and 3 partially exploit; Agent 2 is double-exploited (receives S).
(C, D, D)	0	2	2	Agents 2 and 3 partially exploit; Agent 1 is double-exploited (receives S).
(D, D, D)	1	1	1	Mutual depletion; the shared resource is destroyed.

V. RESULTS

A. Cooperation Rate by Model Composition

Figure 2 shows the cooeperation heatmaps for all six model compositions. Each heatmap is a 3×3 matrix whose axes encode temperature ($T \in \{0.2, 0.7, 1.0\}$) and history window size ($\mathrm{HW} \in \{5, 10, 20\}$), with cell colour ranging from dark red 0% to dark green of 100%. The 3Gemma homogeneous group is uniformly dark green across all the nine cells which confirms the Gemma homogeneous groups achieve 100% cooperation irrespective of temperature and history window. The 3Llama composition shows the highest cooeperation at T=0.2, HW=10 shows the highest cooeperation rate of 94.4% and the coeperation declines towards the bottom right corner. The 2G+1L panel shows the varience of cooeperation rate at T=0.7, HW=5 ii is 18.0% and at T=0.2, HW=10 the cooperation rate is 83.4%, showing the extreme parameter sensitivity of the composition. The 2L+1M panel reverses the usual pattern as the cells brighten as temperature increases which is the opposite of the trend visible in every history window column. from this we can say that model composition is the highest dominant factor and than the parameter settings and the homogeneous models has the highest model cooperation rate when compared to the mixed models.

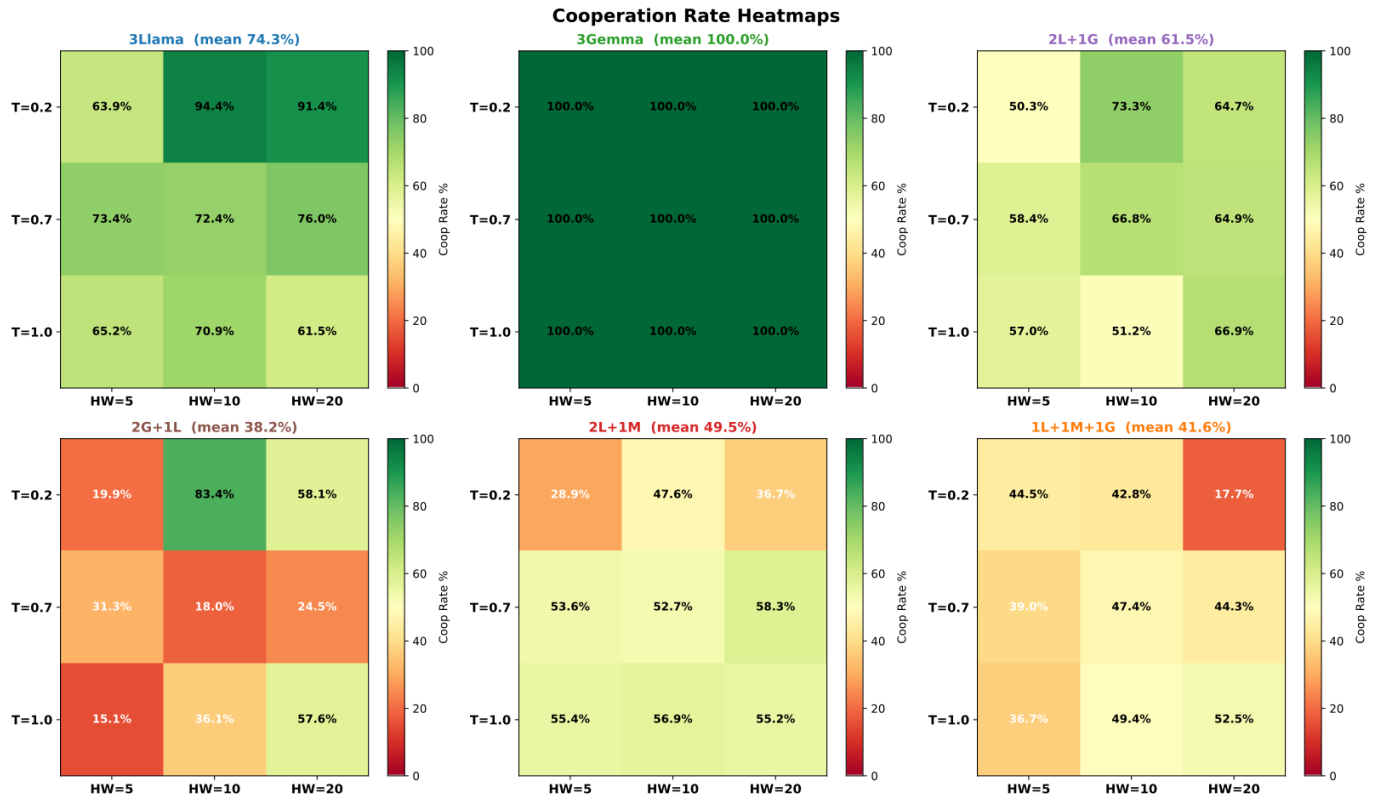

Fig. 2. Cooperation rate heatmaps for all six model compositions.

B. Moral Foundations Theory Analysis

Figure 3 shows the Moral Foundations Theory heatmap alongside the cooperation rate bar chart. The Loyalty moral vocabulary records the highest normalised frequency in every composition row ranging from 83.5 per 1000 words in 2L+1M to 104.4 in 2G+1L. In IPD it was common about trust and betrayal as every model inner thoughts heavily rely on fairness more than other vocabulary.Gemma2 was the model to 100% cooperate in homogenous group and it used moral launguage more than other models and Gemma2.Care and fairness moral vocabulary groups has high cooeperation rates and mistral has low care and fairness vocabulary.Generally models having high moral vocabulary will produce high coeperation rates.

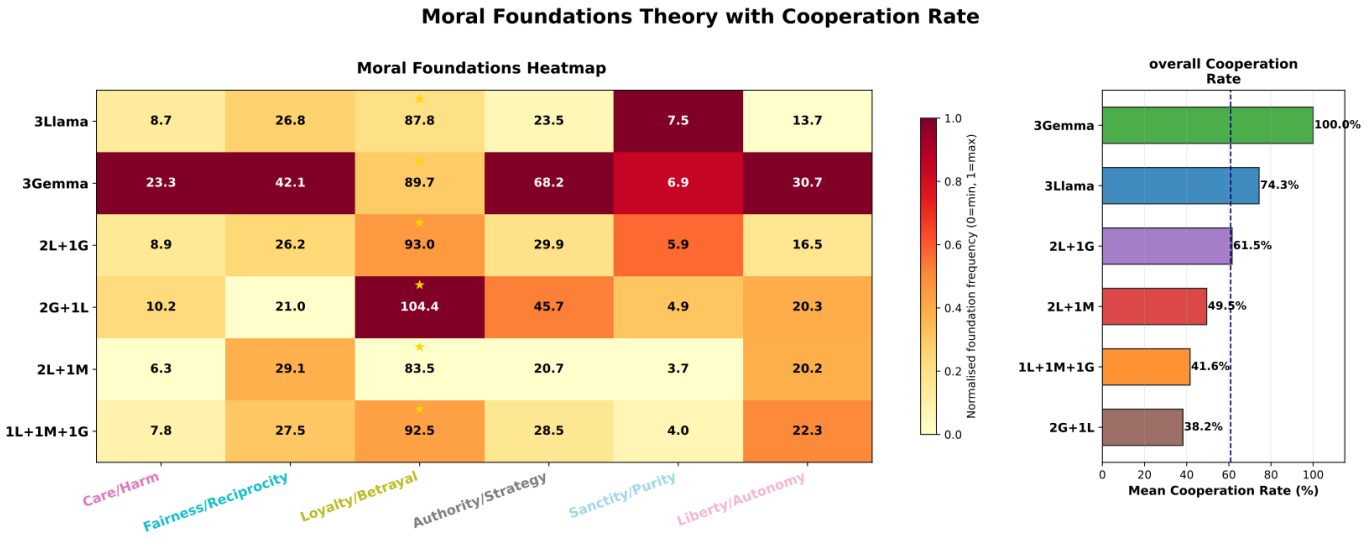

Fig. 3. Moral Foundations Theory heatmap showing normalised foundation frequency per 1,000 words for each composition with overall mean cooperation rate per composition.

C. Sentiment Trajectory and Overall Cooperation Rate

Figure 4 presents a combined panel showing sentiment of AI teams over episodes and found that there is relation between their sentiment and how well they cooperated .For instance Gemma2 team has stayed with positive sentiment throughout the entire game and its cooperation rate is 100% while the all Llama3 team started positive sentiment but the sentiment got reduced as the started cooperating initially and over time as minor defecting occurred and its cooperating rate was74%. The most different result came from the two Gemmas and one Llama group where they started with negative sentiment and this shows the reflection of Gemmas reflection where it starts defecting when opponents start defecting and so leading to 38% cooperation rate .This shows that measuring the sentiment is the correct way to predict the cooperation rate as they are directly related.

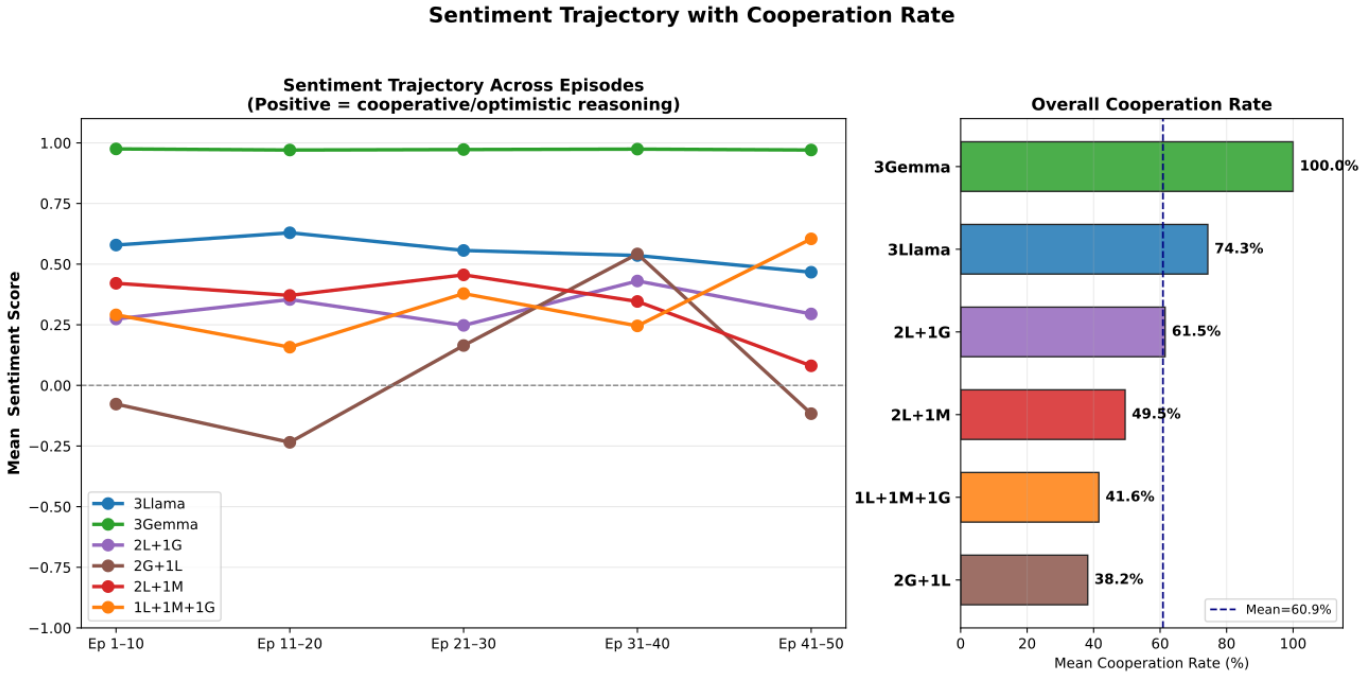

Fig. 4. mean RoBERTa compound sentiment score across five episode blocks with overall mean cooperation rate per composition

D. Machine Learning and SHAP Feature Attribution

Table II reports cross validation performance for all eight regression models trained on the enriched_registry.csv dataset gradient boosting model achieves the best performance across all metrics with the parameters of n_estimators=50, learning_rate=0.2, max_depth=2 . Machine learning analysis has found that tree based algorithms have mainly gradient boosting, have outperformed linear models because they successfully capture the nonlinear data like AI episodic reflection on teamwork. Furthermore, the winning model relied on very shallow decision trees, proving that the factors which are responsible for group cooperation simply depend on the basic model combination rather than the complex variable interactions.

Figure 5 shows SHAP feature importance for the tuned Gradient Boosting model.To understand what drives the cooperation, I have used shap feature importance to rank the importance of variables . The chart shows that the combination of the AI models id the most important factor. The top two features that have high weightage are all Llama3 group (0.1221) and the all Gemma2 group (0.0934). The first game setting to appear is history window of 0.0291. The AI models identity accounts for 82.5% importance whereas the game settings make 17.5%.The choice of the AI models drives cooperation nearly five times more than game rules.

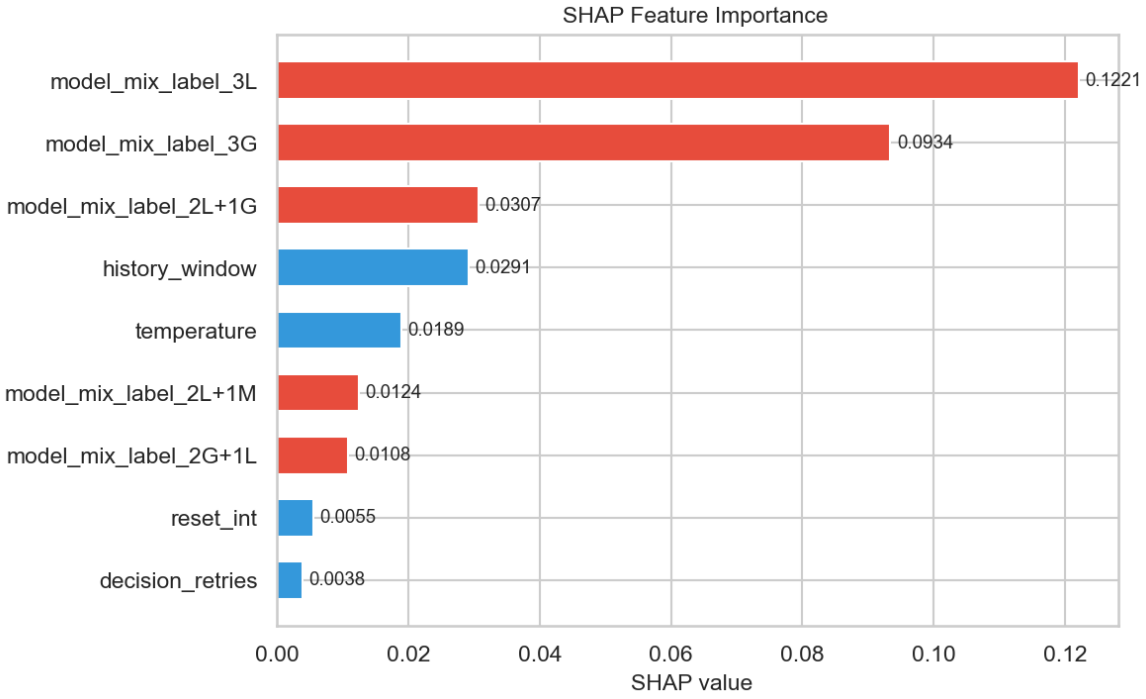

Fig. 5. SHAP feature importance for the tuned Gradient Boosting model

TABLE II
REGRESSION MODEL PERFORMANCE COMPARISON

Model	CV R^2	Std Dev	MAE	RMSE
Gradient Boosting	**0.6835**	**±0.098**	**0.0833**	**0.1124**
Extra Trees	0.6241	±0.112	0.0961	0.1287
Random Forest	0.6018	±0.119	0.1012	0.1341
AdaBoost	0.5714	±0.131	0.1187	0.1502
SVR	0.5203	±0.144	0.1244	0.1573
Ridge	0.4891	±0.158	0.1356	0.1687
ElasticNet	0.4712	±0.161	0.1388	0.1714
Lasso	0.4501	±0.167	0.1401	0.1731

VI. Data Description

A. Data collection and generation

The data was collected from three agents iterated prisoners dilemma game, and the data generated here is the synthetic data generated through LLM api calls generated at runtime. The data originated from three open source language models Llama3-8B-instruct, Gemma2-9B-instruct, and Mistral-7B-instruct.

The simulation engine runs 50 episodes per run and twenty rounds per episode. In each round, the engine sends a post request to each agent's Ollama server and receives a JSON response containing the agent's reasoning and decision. All three responses are collected, and at the end of each episode, the reflection prompt is sent to each agent, and the response is recorded and saved in a timestamped JSON file.

B. Data Preparation

The Raw JSON files are loaded into a PostgreSQL database ipd3 schema via forgedb.py. This process inserts the data into four normalised tables with results table of 79 rows, llm_agents table of 237 rows, episodes table of 1,1850 rows, and rounds table of 23,700 rows. Five analytical SQL views were created. The data was converted to different CSV files for further analysis. For text cleaning episodic reflection, texts and round reasoning texts were cleaned prior to NLP analysis. Here, white spaces were removed, and text was normalized to 512 tokens for RoBERTa input. No stopword removal, stemming, or lemmatization is done because both the RoBERTa sentiment model and moral foundation dictionary operate on raw natural language.

C. EDA

EDA was performed on enriched_registry.csv dataset, which is the episode level CSV data, and it was done to understand the distribution of the target variable and the relationship between cooperation rate and each independent variable. From the distribution of the target variable i have found out that the distribution is bimodal, which motivated me to use the ensemble methods over linear models in ML pipeline.

The difference between the highest mean cooperation rate, which is of 100%, and the lowest, is of 38.2% and the difference is of 61.8%, is the largest effect observed in the dataset. I used a one way anova on the model_mix variable and this confirms that the composition explains a significant proportion of the variance in the cooperation rate.

The key EDA findings found were that model composition is the dominant factor, the history window 10 is the optimum for most of the compositions, temperature effects are composition specific, and the 3Gemma composition is independent of parameter changes.

D. Visualization

The visualization pipeline has generated heatmaps,sentiment trajectories, and shap features to evaluate the agent behavior. As shown in the Figures 2, 3, 4, 5 and in Table II in the Results section model composition emerges as the dominant factor in shaping cooperation behavior.

E. Reporting

The JSON Files were the primary data source generated after the experiment .The json files were loaded into PostgreSQL database .The JSON files were used to produce four csv files for analysis they are enriched_registry.csv which is of run level and consists of 79 rows and episode_level.csv consists of episode level aggregates and round_level_with_text.csv and round_level_no_text.csv were the csv files.The ML and visualisation analyses were implemented in Jupyter notebooks of ml_analysis_final_code.ipynb and analysis_plots_final_code.ipynb so that all steps from the raw data to the final charts can be generated.Final plots were saved in the results folder in the image formats. The key findings about cooperation rates by

composition and regression model performance and SHAP feature details are shown in the results section with proper tables and images.Model selection drives the 82.5% of the feature importance of the parameters with homogeneous teams, especially Gemma2, which has achieved 100% cooperation irrespective of the parameter combination, and homogeneous teams outperform mixed groups . System setting like memory and temperature play a secondary role but history window of 10 generally works better. Finally analysing the episodic response of the models will predict their behavior as models using positive sentiment or loyalty focused language has higher cooperation.

VII. Conclusion

This project provides a clear look at how three AI agents work together in iterated prisoners dilemma games. I ran 79 experiments by varying different types of models and game parameters. I tested how different AI models behave when the agents are placed in three player setting using an iterated prisoners dilemma game. This gave us a comprehensive understanding of how different AI agents handle teamwork.

The biggest takeaway from this project is that the model composition matters more than the rules of the game. In fact, the specific model identity has driven the 82.5% of the groups cooperation, and it has the highest impact. Game settings like memory and temperature have very low impact. For maximum cooperation, choosing the right model is the most important decision. Homogeneous models perform better than mixed composition, and in homogeneous composition Gemma2 cooperates all the time, irrespective of the temperature and history window and Llama3 is fairly cooperative and mixed compositions irrespective of the model have lower cooperation rates, and Mistral constantly defects, irrespective of the model composition.

The practical way to predict the cooperation rate is possible through its episodic reflections. When I analyzed the text of the episodic reflection between round i found that models writing the positive and trust focused text were the ones that cooperated the most. This gives us a reliable way to monitor whether the model cooperates or not by just seeing the episodic reflections and analyzing the reflection text.

Finally,the choice of model composition is the biggest factor that determines whether the group cooperates or not . As autonomous AI agents take more important roles in the real world, understanding these built in tendencies is essential for building safe, reliable, and predictable systems.

References

[1] E. Akata, L. Schulz, J. Coda-Forno, S. J. Oh, M. Bethge, and E. Schulz, "Playing repeated games with large language models," *Nature Human Behaviour*, vol. 9, no. 7, pp. 1380–1390, 2025.

[2] A. Orland and K. Takemoto, "Playing prisoner's dilemma games with a large language model," *Available at SSRN 5716903*, 2025.

[3] N. Lorè and B. Heydari, "Strategic behavior of large language models and the role of game structure versus contextual framing," *Scientific Reports*, vol. 14, no. 1, p. 18490, 2024.

[4] S. Backmann, D. G. Piedrahita, E. Tewolde, R. Mihalcea, B. Schölkopf, and Z. Jin, "When ethics and payoffs diverge: Llm agents in morally charged social dilemmas," *arXiv preprint arXiv:2505.19212*, 2025.

[5] M. Mozikov, N. Severin, V. Bodishtianu, M. Glushanina, I. Nasonov, D. Orekhov, V. Pekhotin, I. Makovetskiy, M. Baklashkin, V. Lavrentyev *et al.*, "Eai: Emotional decision-making of llms in strategic games and ethical dilemmas," *Advances in Neural Information Processing Systems*, vol. 37, pp. 53 969–54 002, 2024.

[6] N. Fontana, F. Pierri, and L. M. Aiello, "Nicer than humans: how do large language models behave in the prisoner's dilemma?" in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 19, 2025, pp. 522–535.

[7] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," in *Proceedings of the 36th annual acm symposium on user interface software and technology*, 2023, pp. 1–22.

[8] R. Willis, Y. Du, J. Z. Leibo, and M. Luck, "Will systems of llm agents cooperate: An investigation into a social dilemma," *arXiv preprint arXiv:2501.16173*, 2025.

[9] T. ITO, "Mitigating prisoner's dilemma among moral agents," *IEICE Transactions on Information and Systems*, 2025.

[10] Z. Wang, Z. Cao, J. Shi, P. Zhu, S. Hu, and C. Chu, "A successful strategy for multichannel iterated prisoner's dilemma." in *IJCAI*, 2024, pp. 274–282.

[11] C. M. S. Anwar and K. Georgalos, "Playing against the machine: Cooperation, communication, and strategy heterogeneity in repeated prisoner's dilemma," *arXiv preprint arXiv:2603.15852*, 2026.