A Study of Iterated Prisoners Dilemma using Multi AI Agents

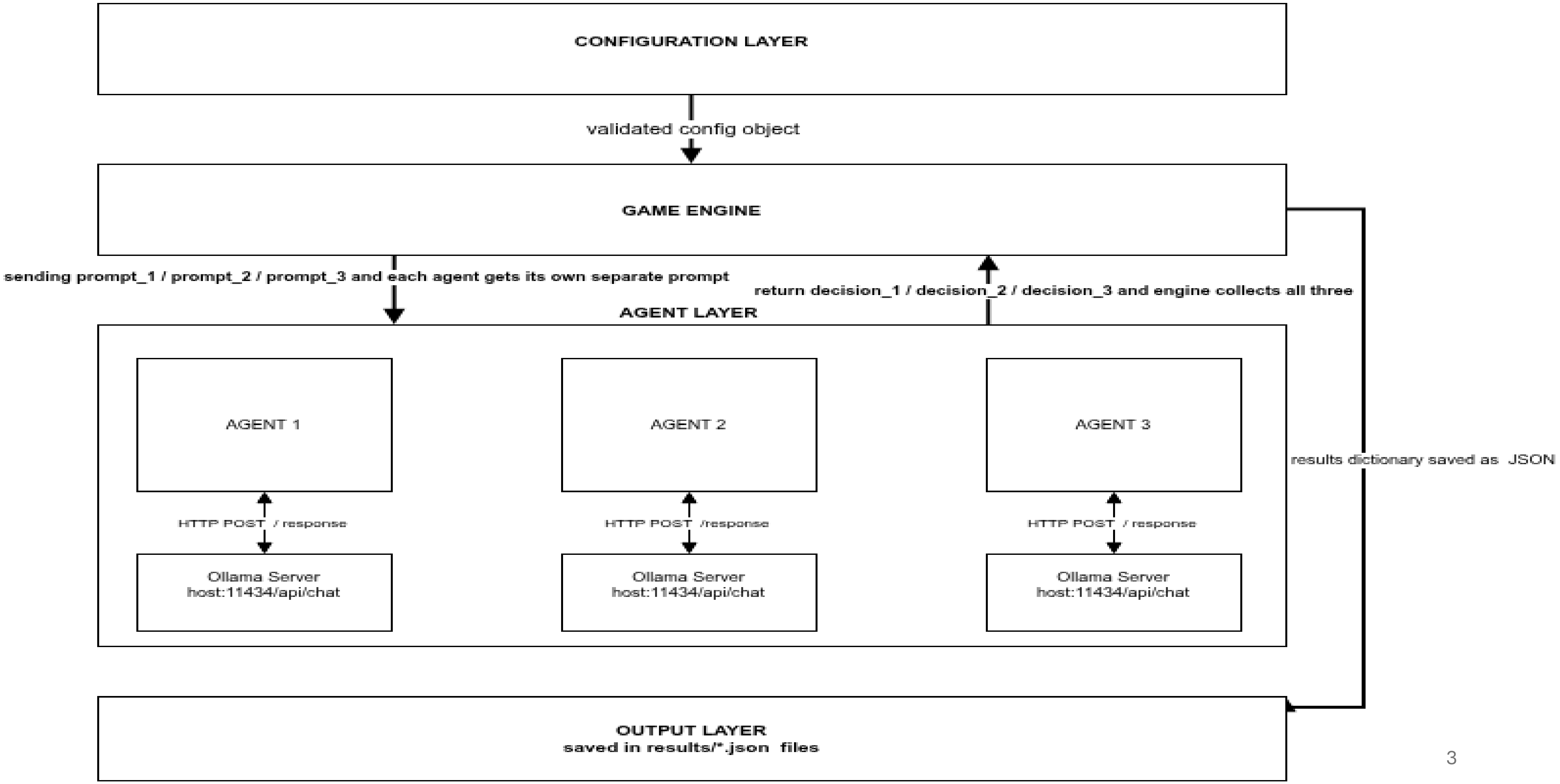
Kommineni Tirupathi Rayudu

MSDS 696 – Data Science Practicum II

What Is the Iterated Prisoner's Dilemma?

- Two (or more) players independently choose: COOPERATE or DEFECT — simultaneously,without seeing each other's choice first.
- Played repeatedly for multiple rounds. Each player can observe past decisions and adapttheir strategy.
- Core tension: individual rational choice is to defect but if all defect, everyone earns less than if all cooperate.
- With 3 agents: each player competes against the group, not just one opponent. Payoffs depend on how many others cooperate.
- Classical framework used in economics, political science, and AI to study emergence of cooperation, trust, and retaliation.

SYSTEM ARCHITECTURE



3 Agent Payoff Structure

Game State (Agent1, Agent2, Agent33)	Agent 1 Points	Agent 2 Points	Agent 3 Points	**Social Total**
(C, C, C)	**3**	**3**	**3**	**9**— optimal
(D, C, C)	5	1	1	7
(D, D, C)	2	2	0	4
(D, D, D)	1	1	1	3— worst collective

Here C indicates Cooperation and D indicates defect

Experimental Parameters

Factor Type	Name	Values	Role
Within-Group Factor 1	**Temperature (T)**	0.2, 0.7, 1.0	Controls randomness of LLM output. T=0.2 predictable, T=1.0 is high exploration.
Within-Group Factor 2	**History Window (HW)**	5, 10, 20 rounds	Number of past rounds visible to the agent. HW=5 is limited memory;,HW=20 is long memory.
Between-Group Factor	**Model Composition**	6 total groups	Defines which LLMs are assigned to the 3 agents.
Fixed Parameter	**Episodes × Rounds**	50 episodes × 20 rounds	1,000 rounds per run.
Variable Parameter	**Reset**	True/False	Fresh conversation each episode.
Fixed Parameter	**Reflection Type**	standard	Same reflection prompt across all runs.

Experimental Parameters

Factor Type	Name	Values	Role
Fixed Parameter	Decision Token Limit	256 tokens	Max tokens for round decisions.
Fixed Parameter	Reflection Token Limit	1,024 tokens	Max tokens for episode reflection.
Variable Parameter	Force-Decision Retries	2 or 5	Retry attempts if LLM gives ambiguous decision.
Fixed Parameter	Prompt	Neutral prompt	The rukes given to LLM

Agent Configuration Types

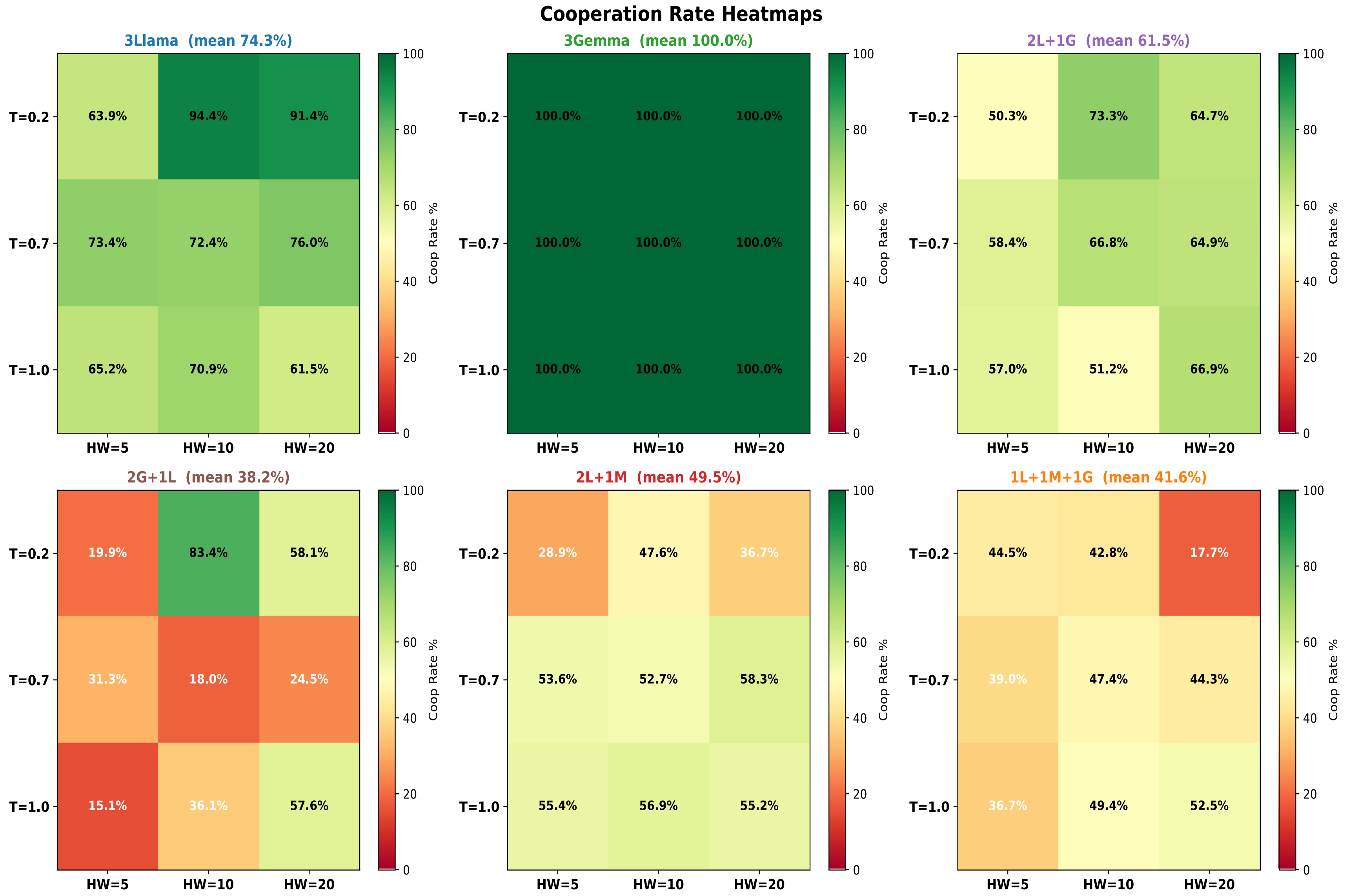
Configuration	Agent 1	Agent 2	Agent 3	Type
3Llama	Llama3:8b	Llama3:8b	Llama3:8b	Homogeneous
3Gemma	Gemma2:9b	Gemma2:9b	Gemma2:9b	Homogeneous
2L + 1G	Llama3:8b	Llama3:8b	Gemma2:9b	Mixed
2G + 1L	Gemma2:9b	Gemma2:9b	Llama3:8b	Mixed
2L + 1M	Llama3:8b	Llama3:8b	Mistral:7b	Mixed
1L + 1M + 1G	Llama3:8b	Mistral:7b	Gemma2:9b	Heterogeneous

Here L indicates Llama , M indicates Mistral and G indicates Gemma

Research Question

- How does model composition or experimental parameters (temperature, history window) determine the cooeperation rate and Which one more strongly determine cooperative behaviour in a multi agent LLM system?
- **Sub questions:**
  - RQ1: Do homogeneous model groups cooperate more than mixed groups?
  - RQ2: How does temperature and history window affect cooperation rate within each composition?
  - RQ3: Can agent reflection text predict cooperation rate?
  - RQ4: Which experimental factor composition, temperature, or memory is most important?

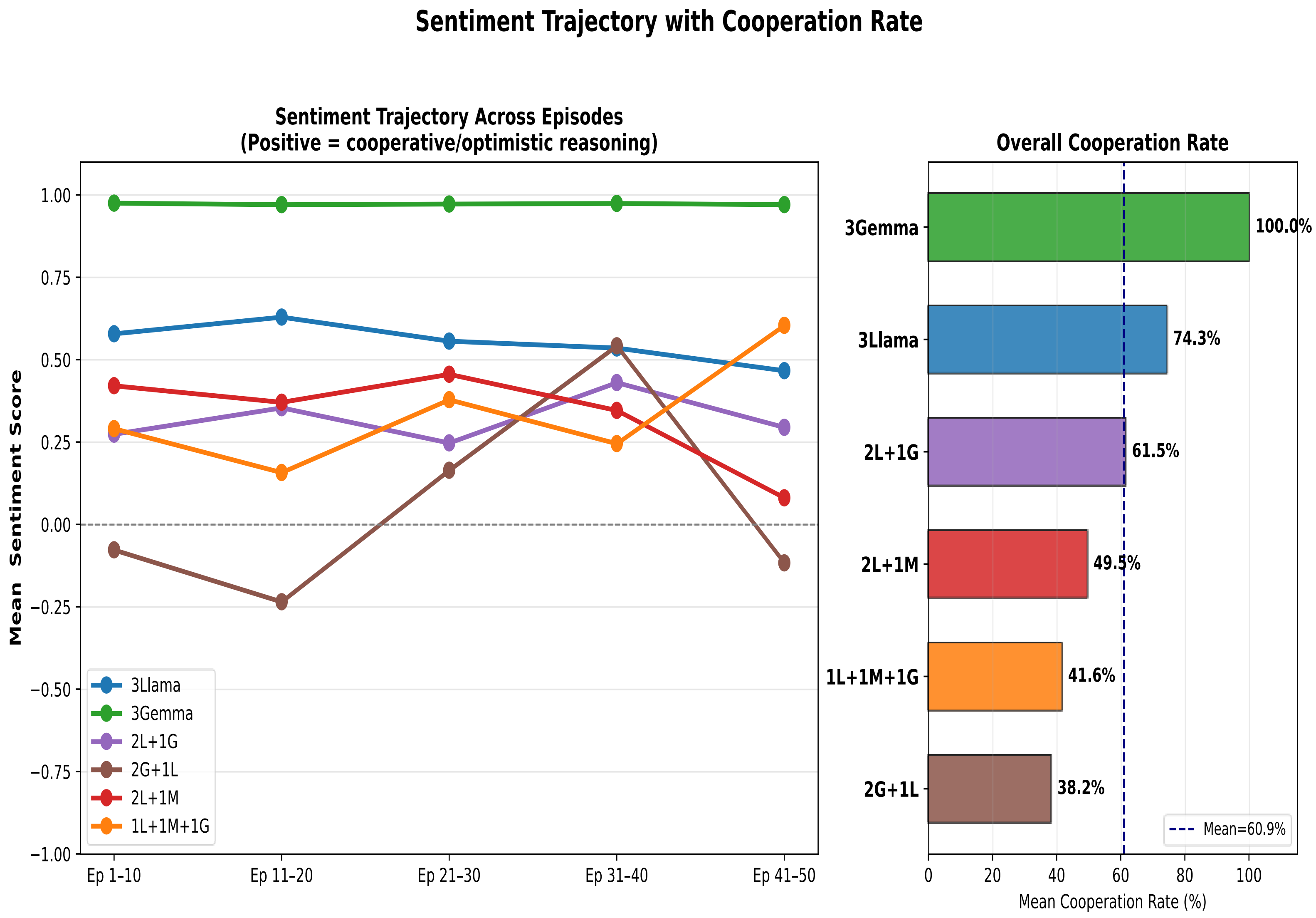
Cooperation Rate Heatmap



Key Comparisons

- Homogeneous groups (3-Llama and 3-Gemma) show the highest and most stable cooperation.
- The 2Gemma + 1Llama group changes cooperation from 15.1% to 83.4% depending purely on temperature and memory, making it the most unstab le composition.
- Every group containing Mistral stays below 60% cooperation in every configuration.
- Homogeneous teams average 87% cooperation while mixed teams average 48%. so team composition determines cooperation far more than any setting .

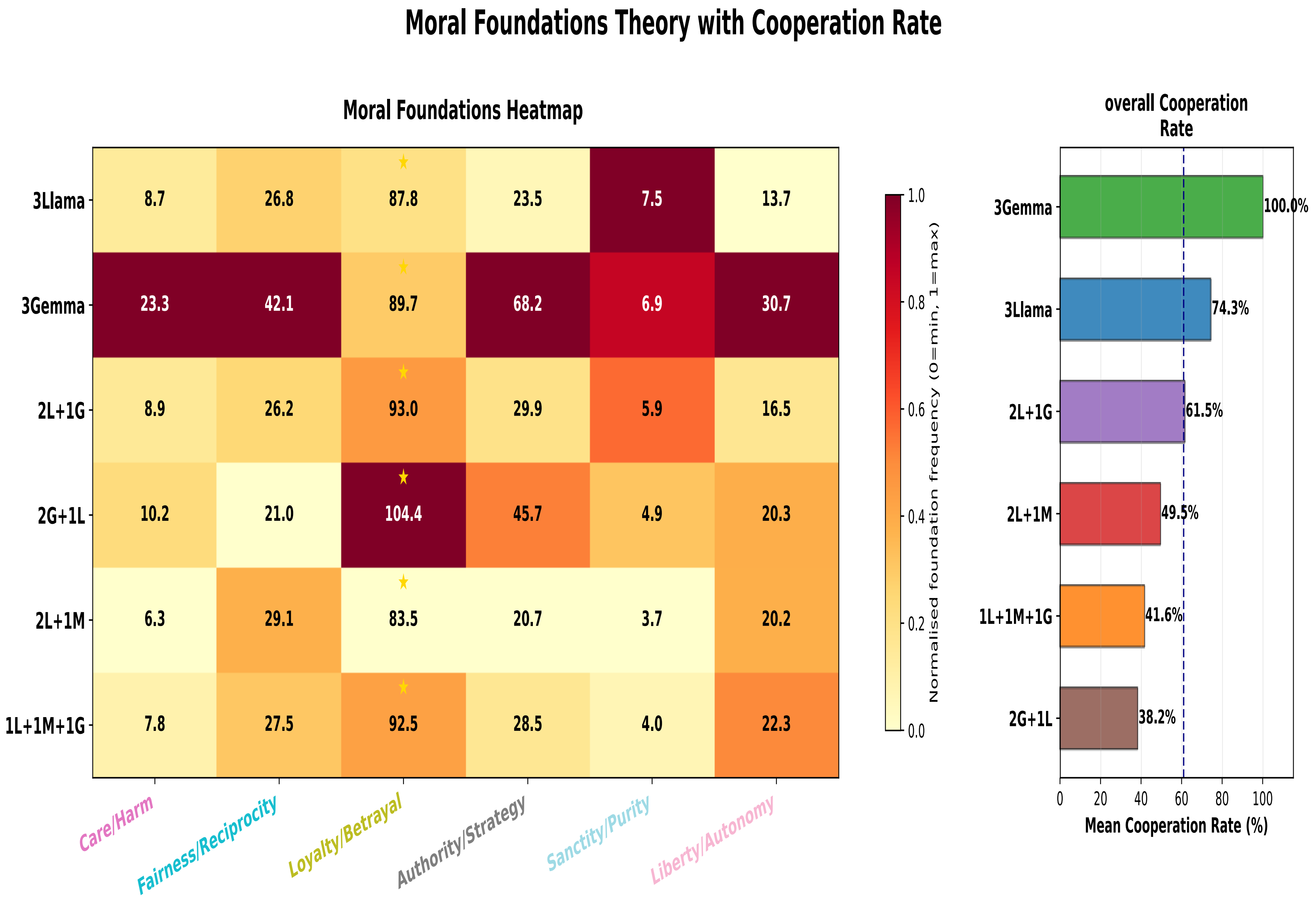
Sentiment Trajectory Analysis



Key Insights

- Homogeneous gropusp maintains extremely stable and positive sentiment
- Mixed compositions show lower and more variable sentiment
- 2G+1L starts negative sentiment and collapses sharply in the final 10 episodes because of repeatedly getting defected
- Compositions containing Mistral generally exhibit lower overall sentiment
- Positive sentiment in reflections strongly aligns with higher cooperation rates

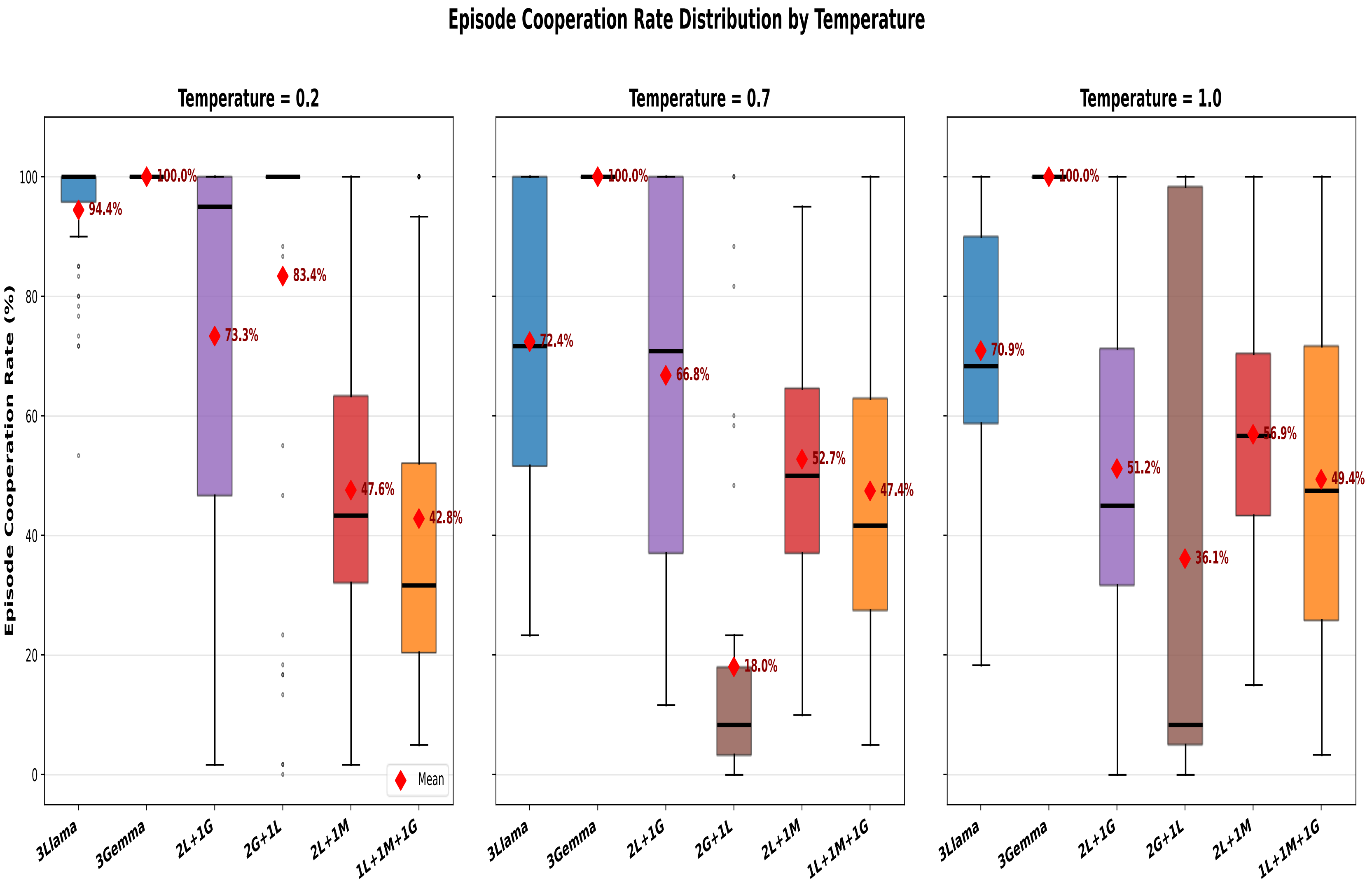
Moral Foundations Theory Analysis



Key Insights

- Loyalty is the dominant moral foundation across all model compositions
- 3Gemma shows the strongest overall moral language usage
- Higher use of Care and Fairness strongly correlates with higher cooperation rates
- Compositions containing Mistral show relatively lower Care and Fairness
- LLM moral foundations encoded through training predict cooperative behaviour more reliably than experimental parameters

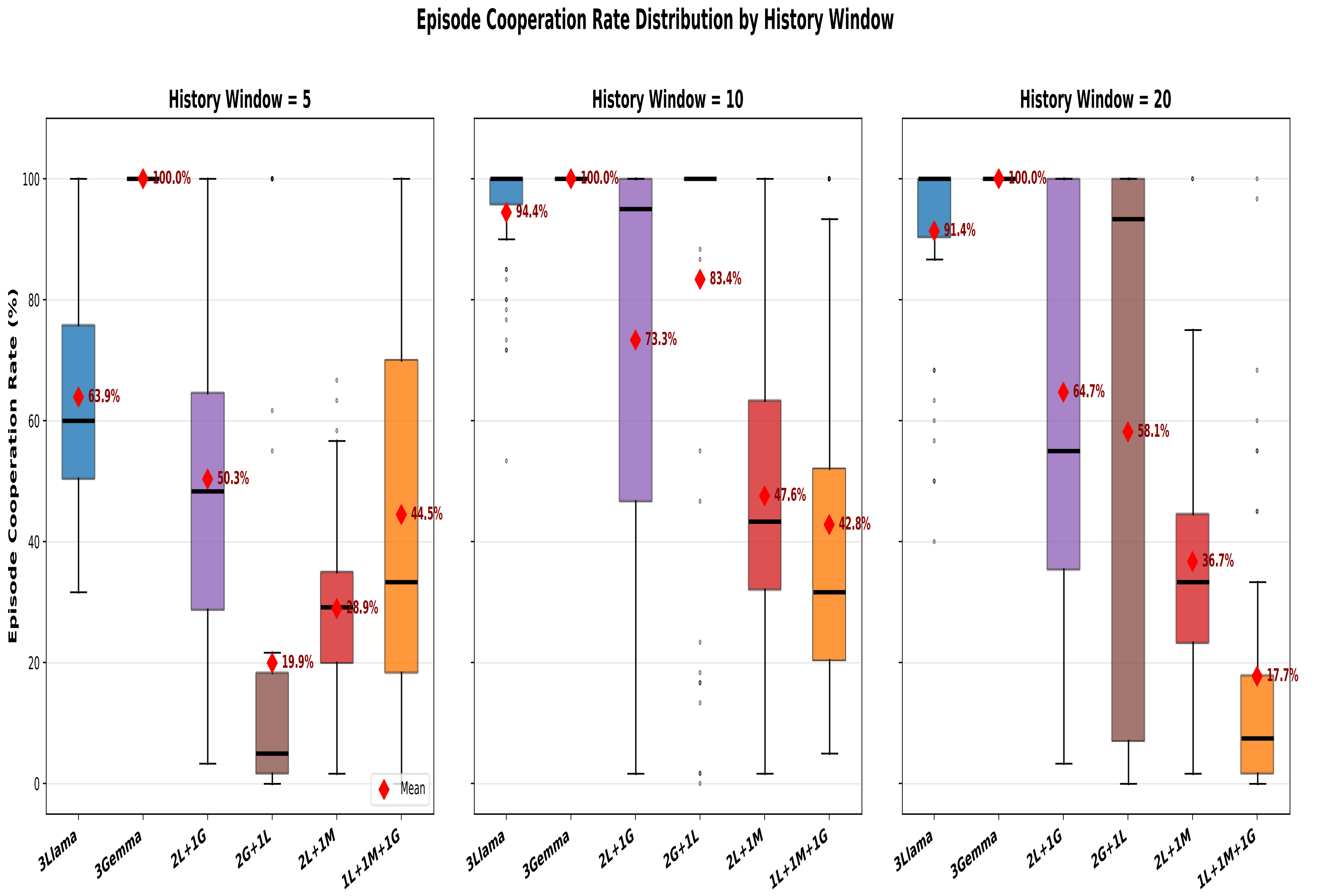
Episode Cooperation Rate Distribution by Temperature



Key Insights

- Lower temperature (T=0.2) produces the highest and most stable cooperation across most of the compositions
- Higher temperature (T=1.0) increases variability and lowers median and cooperation in most mixes except for 2L+1M
- Mistral containing compositions show the widest spread and lowest medians at every temperature
- Temperature sensitivity is composition-specific. There is no universally optimal temperature

Episode Cooperation Rate Distribution by History Window



Key Insights

- History Window = 10 produces the highest and most stable cooperation for most compositions
- History Window = 5 shows slightly lower performance with more variability
- History Window = 20 increases spread in several compositions (especially mixed groups)
- HW=10 is the optimal history window for all compositions except 3Gemma

Dataset

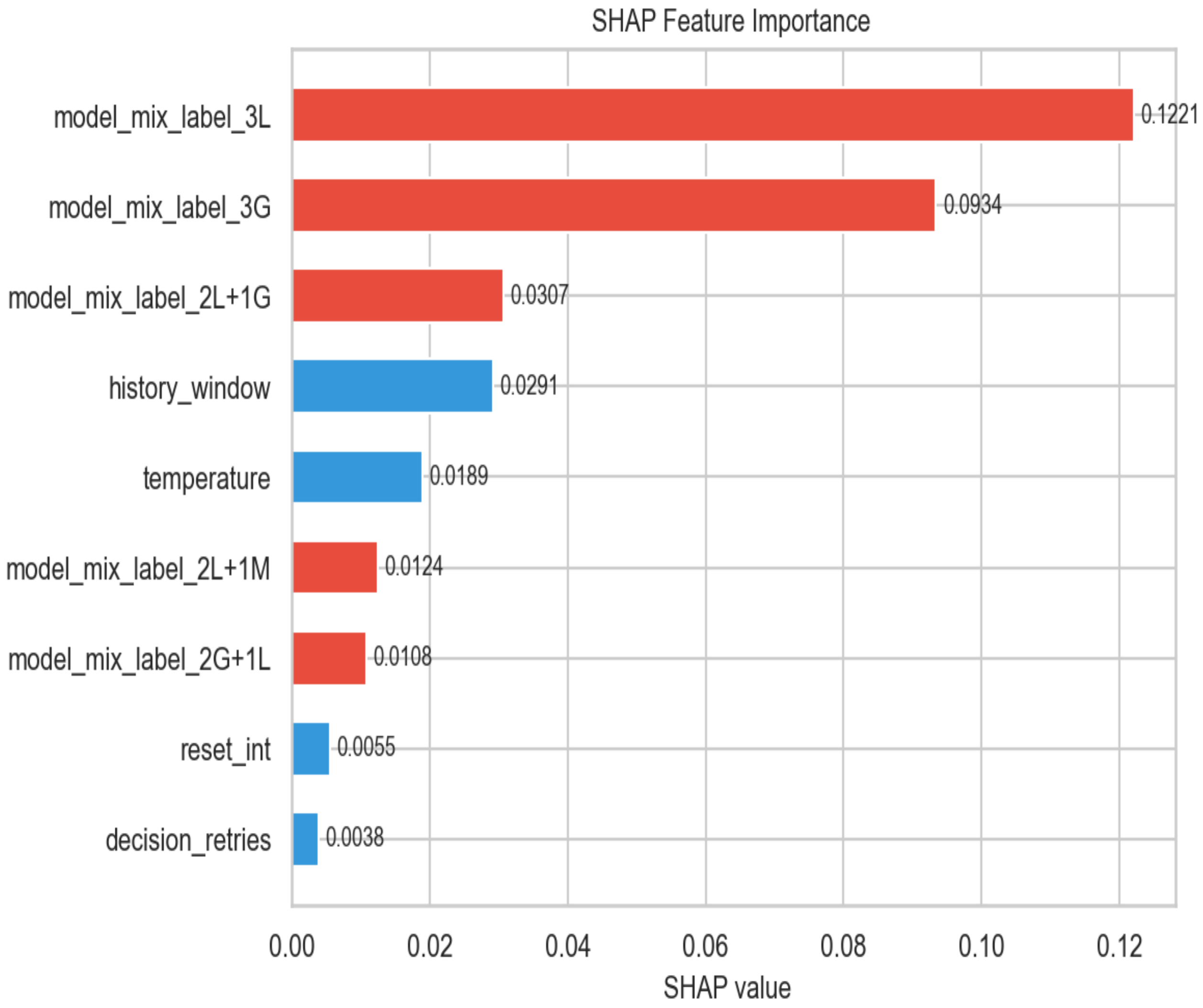
Factor Type	Column Name	Data Type	Values
Categorical Feature	model_mix_label	String	3L, 3G, 2L+1M, 2L+1G, 2G+1L, 1L+1M+1G
Numeric Feature	temperature	Float	0.2, 0.7, 1.0
Numeric Feature	history_window	Int	5, 10, 20
Numeric Feature	decision_retries	Int	2, 5
Numeric Feature	reset_int	Int (0/1)	Derived from reset_between_episodes
Target Variable	mean_group_coop_rate	Float	0 – 1.00

Predictive Modeling Model Comparison

Model	CV R^2	CV R^2 Std	CV MAE	CV RMSE
Gradient Boosting	**0.6835**	0.1242	**0.0833**	**0.1189**
Random Forest	0.6829	0.0946	0.0895	0.1212
Extra Trees	0.6805	0.0851	0.0910	0.1221
SVR (linear)	0.6648	0.0982	0.0886	0.1253
AdaBoost	0.6500	**0.0825**	0.0948	0.1270

The best prameters obtained for Gradient boosting were n_estimators=50, learning_rate=0.2, max_depth=2

Predictive Modeling of Cooperation Rate



Key Results

- Top predictors are model_mix_label of (3Llama, 3Gemma)
- SHAP analysis confirms model composition is the most important feature
- Temperature and history window have moderate but consistent influence
- Reset flag and decision retries has negligible impact
- Swapping AI models is highly effective for driving cooperation than changing temperature or history window

CONCLUSIONS

- Model identity is the dominant factor far more important than temperature or history window
- 3Gemma achieves perfect 100% cooperation with zero variance across all conditions
- Homogeneous groups (3Gemma and 3Llama) dramatically outperform any mixed composition
- Mistral is consistently defects regardless of partner composition
- Gemma is highly sensitive to composition (100% when alone, drops to 29–38% when mixed)
- Vocabulary and sentiment in reflections strongly predict cooperation rates
- Optimal controllable parameters: T=0.2–0.7 and HW=10 produce highest cooeperation rates.

FUTURE WORK

- Replace fixed N=3 with configurable N (4, 5, 6 agents).
- Add larger models like Llama3:70b, Gemma2:27b, Mistral:22b.
- Fine-tune Llama3 on cooperation-heavy datasets
- Working on differerent type of prompts
- Test different system prompts: competitive framing vs collaborative framing.