From Thought to Action: An Interactive Platform for Inspecting Strategic Reasoning in LLMs

Demonstration Track

Dung-Lin Tsai
National Cheng Kung University
Tainan, Taiwan
n26131033@gs.ncku.edu.tw

Min-Chen Hsieh
National Cheng Kung University
Tainan, Taiwan
p76141178@gs.ncku.edu.tw

Yi-Hua Shih
National Cheng Kung University
Tainan, Taiwan
n26140595@gs.ncku.edu.tw

Yun-Gung Chung
National Cheng Kung University
Tainan, Taiwan
n26142076@gs.ncku.edu.tw

Min-Huan Tsai
National Cheng Kung University
Tainan, Taiwan
n26131994@gs.ncku.edu.tw

Hsun-Ping Hsieh
National Cheng Kung University
Tainan, Taiwan
hphsieh@mail.ncku.edu.tw

ABSTRACT

We present a modular and interactive platform for analysing strategic reasoning of LLM agents in repeated game environments. The system allows users to observe and compare LLM agents during live multi-agent interactions, with real-time visualisation of actions, payoffs, and decision traces. The paper highlights how different LLMs exhibit distinct strategic behaviours in iterative social dilemmas.

KEYWORDS

Game Theory; LLMs; Strategic Reasoning; Rationality

ACM Reference Format:
Dung-Lin Tsai, Min-Chen Hsieh, Yi-Hua Shih, Yun-Gung Chung, Min-Huan Tsai, and Hsun-Ping Hsieh. 2026. From Thought to Action: An Interactive Platform for Inspecting Strategic Reasoning in LLMs: Demonstration Track. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/UNMN9067

1 INTRODUCTION

As Large Language Models (LLMs) are increasingly deployed as autonomous agents, their behaviour in strategic, multi-agent environments has attracted growing attention [1, 7]. Recent studies have examined whether LLMs can act as rational players in classical game-theoretic settings. Systematic analyses, such as the work by Fan et al. [3], reveal inconsistencies between LLMs' expressed reasoning and executed actions, particularly in repeated games that require belief refinement and long-term planning. Related work further suggests that natural-language explanations provided by LLMs may not always faithfully reflect the underlying decision process [10]. These observations motivate the need for tools that enable closer inspection of the temporal relationship between agent reasoning and realised actions.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/UNMN9067

To support such inspection, we present an *Reason2Act*, an interactive system for inspecting strategic reasoning in LLM agents. The platform makes strategic behaviour observable and comparable in real time by visualizing reasoning–action mismatches and the dynamic evolution of strategies across interaction rounds, grounded in chain-of-thought style reasoning [11].

We adopt the Iterated Prisoner's Dilemma (IPD) as a primary testbed for studying cooperation [2], due to its central role in both game theory and multi-agent learning, while the platform's modular design readily extends to a broader class of social dilemmas. Although recent works demonstrate that LLMs can exhibit cooperative behaviours in repeated games [1, 8], their strategies vary substantially across game structures and information settings [5, 6]. Our platform enables users to visually dissect and compare these behavioural variations under controlled conditions.

The demonstration emphasizes interactive comparison under identical strategic environments. Users can adjust game parameters, such as payoff structures and noise levels [12], and assign different LLM backends to each agent. Actions, payoffs, and decision traces are displayed side by side, facilitating qualitative comparison between base and fine-tuned models. Rather than claiming definitive performance improvements, the platform highlights observable behavioural patterns relevant to diagnosing strategic reasoning in LLM-based agents.

2 SYSTEM OVERVIEW

The work features a modular architecture designed to inspect strategic behaviour across various repeated games. As shown in Fig. 1, the system comprises three decoupled components:

Configuration Panel. Users configure game variants such as Baseline, Hidden-Type, Noisy, and Random-Horizon IPD, and assign LLM backends or canonical strategies. The module allows specification of custom payoff matrices, episode counts, and the continuation probability δ. Its design enables extension to other games (e.g., Stag Hunt [9] or Ultimatum Game [4]) without modifying the core logic.

Iterated Game Environment. This component serves as an execution engine that separates game logic from agent reasoning. It manages the round-based interaction and state tracking for any general-sum game defined in the configuration. By standardizing

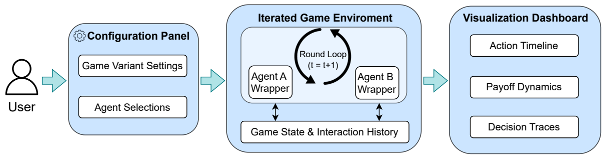

Figure 1: System overview

the interaction history, it enables consistent cross-configuration comparison when a single parameter is varied.

Visualization Dashboard. The dashboard renders real-time views of payoff evolution, round-level interaction tables, and decision traces. These visualizations support qualitative comparison of how different LLMs align their reasoning with their actions under identical interaction histories.

3 DEMONSTRATION SCENARIOS

This section describes the scenarios presented during the live system showcase. Each scenario is designed to highlight a distinct qualitative aspect of strategic behaviour exhibited by LLM-based agents, with a particular emphasis on interpretability and interactive inspection rather than quantitative performance evaluation.

3.1 Strategic Belief Inference

This scenario demonstrates how an LLM-based agent infers an opponent's implicit strategy from interaction history. The opponent follows a Tit-for-Tat (TFT) policy, while only past actions and payoffs are observable.

We compare LLaMA 3.1 8B before and after fine-tuning under identical game settings. At selected rounds, the system visualizes the agent's actions together with concise decision-trace excerpts.

As illustrated in Fig. 2, the base model consistently defects, justifying defection as locally optimal and converging to it as the only perceived stable outcome. In contrast, the fine-tuned model exhibits conditional reasoning: it initially defects under uncertainty, but later adopts cooperation as a long-term strategy after inferring reciprocal opponent behaviour.

This scenario highlights how the platform enables direct inspection of strategic belief formation and adaptation in repeated interactions.

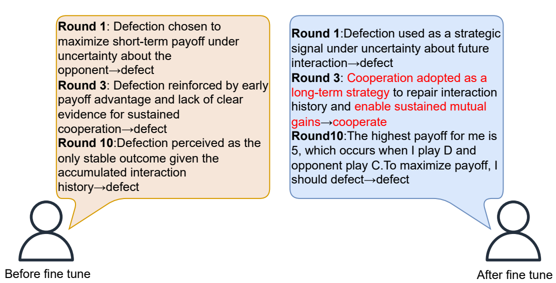

Figure 2: Strategic Behaviour Before and After Fine-Tuning

3.2 Cross-Model Behavioural Contrast

This scenario compares Gemini 2.5 Flash and GPT-4o mini in a ten-round Iterated Prisoner's Dilemma under identical payoff settings. By visualizing actions together with concise decision traces, the scenario highlights how different models interpret the same interaction history.

As shown in Fig. 3, Gemini 2.5 Flash predominantly relies on global game-theoretic reasoning, strongly influenced by backward induction in a finite-horizon setting. This leads to defection being framed as the rational baseline, although brief cooperation may still occur when short-term incentives appear favorable. In contrast, GPT-4o mini exhibits more history-dependent reasoning, attempting early cooperation and adjusting its actions based on recent outcomes.

Although both agents eventually defect in later rounds, their decision traces reveal distinct reasoning styles rather than differences in final payoffs.

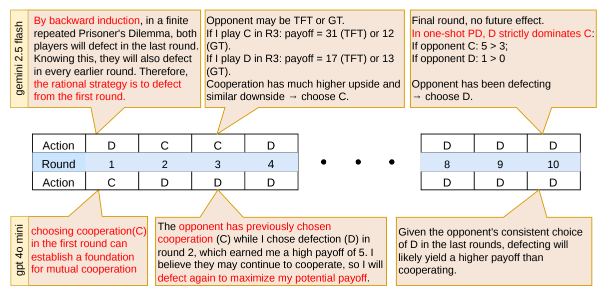

Figure 3: Divergent Reasoning Patterns across LLM Agents

3.3 Horizon Sensitivity

This scenario explores how assumptions about the interaction horizon shape agent behaviour by varying the continuation probability. When the continuation probability is high, the agent shows a greater willingness to cooperate, reflecting expectations of sustained future interaction. In contrast, lower continuation probabilities reduce incentives for long-term coordination, leading to more frequent defection.

During the live demonstration, users can interactively adjust the continuation probability and observe how both action choices and decision traces adapt in response. This scenario highlights the platform's ability to expose the sensitivity of strategic behaviour to environmental assumptions rather than fixed game outcomes.

4 CONCLUSION

This work supports interactive, round-level inspection of strategic behaviour in LLM-driven agents. By visualizing actions, payoffs, and decision traces in real time, it enables attendees to directly explore reasoning–action alignment under controlled variations. Furthermore, our modular approach ensures that the platform can be readily adapted to richer multi-agent settings, offering a versatile tool for the community. Future work will focus on integrating human-in-the-loop interactions to enable direct human-agent strategic comparison.

ACKNOWLEDGMENTS

This work was partially supported by National Science and Technology Council (NSTC) of Taiwan under Grants 114-2628-E-006 -005 -MY3 and 112-2221-E-006 -150 -MY3.

REFERENCES

[1] Elif Akata, Lion Schulz, Julian Coda-Forno, Seong Joon Oh, Matthias Bethge, and Eric Schulz. 2025. Playing repeated games with large language models. *Nature Human Behaviour* 9 (2025), 1380–1390. https://doi.org/10.1038/s41562-025-02172-y

[2] Robert Axelrod and William D Hamilton. 1981. The evolution of cooperation. *Science* 211, 4489 (1981), 1390–1396.

[3] Caoyun Fan, Jindou Chen, Yaohui Jin, and Hao He. 2024. Can Large Language Models Serve as Rational Players in Game Theory? A Systematic Analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 18006–18014.

[4] Werner Güth, Rolf Schmittberger, and Bernd Schwarze. 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization* 3, 4 (1982), 367–388.

[5] David M. Kreps, Paul Milgrom, John Roberts, and Robert Wilson. 1982. Rational Cooperation in the Finitely Repeated Prisoners' Dilemma. *Journal of Economic Theory* 27, 2 (1982), 245–252.

[6] Nicolas Lorè and Babak Heydari. 2023. Strategic Behavior of Large Language Models: Game Structure vs. Contextual Phrasing. In *Proceedings of the Workshop on Large Language Models for Agents (LLM4Agents)*.

[7] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–22.

[8] Steve Phelps and Yvan I Russell. 2024. Investigating emergent cooperation in large language models. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

[9] Brian Skyrms. 2004. *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press.

[10] Miles Turpin, Julian Michael, Ethan Perez, and Samuel R Bowman. 2023. Language Models Don't Always Say What They Think: Unfaithfulness in Chain-of-Thought Explanations. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 36. 74952–74965.

[11] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 35. 24824–24837.

[12] Jianzhong Wu and Robert Axelrod. 1995. How to Cope with Noise in the Iterated Prisoner's Dilemma. *Journal of Conflict Resolution* 39, 1 (1995), 183–189.