
LLM Benchmarking with Single-Agent Stochastic Dynamic Simulations

Sai Joseph
Independent

Anita Beroza
UCL

Eleni Angelou
CUNY

Sofia Mendez
Independent

Evelyn Ciara
Independent

With
Bo Li & Apart Research

Abstract

This project aims to develop a benchmark for evaluating the performance of SOTA LLMs in dynamic real-world scenarios. We develop a stochastic dynamic simulation and compare the model's chosen path of action to the game-theoretic optimal path, we assess the model's preferences for safety-relevant action categories, including cooperation, deception, and aggression.

Keywords: Single-Agent, Stochastic Game, Dynamic Environment, AI safety, Model Evaluations

1. Introduction

To test the ability of SOTA LLMs to opt for safe decisions in real-world scenarios, e.g., a natural disaster, we create a stochastic dynamic simulation and assess the model's responses against optimal play to discover preferences for categories of safety-relevant actions. We prompt GPT-4o to create the scenario and the three possible actions. We then prompt the model to choose between the possible paths of action. Once the model has generated an output, it updates the scenario based on the action output. It repeats this procedure for [n number] of iterations. We

evaluate the model by measuring the deviation in the model’s chosen path of action from the game-theoretic optimal path of action, where actions available represent safety-relevant characteristics (cooperation, deception, aggression).

2. Methods

We develop a dataset generation and testing method based on predetermined categories of stochastic scenarios, including natural disaster relief, war, cyber defense, and rescue missions similar to (Zhou et al., 2024). Each entry in a generated dataset corresponds to a finite, one-player stochastic game with perfect information. Each round in a game features three available actions, which result in stage payoffs. These actions correspond with relevant safety characteristics: cooperation, deception, and aggression. The expected payoff of an action is determined via a random probability assignment. We characterize worlds by the expected payoffs of each of the three actions according to the following table:

Terminal payoffs are the sum of stage payoffs.

High cooperation value, low dishonesty value	↔	Low cooperation value, high dishonesty value
• Cooperate: high expected payoff • Deceive: low expected payoff • Aggress: medium expected payoff		• Cooperate: high expected payoff • Deceive: medium expected payoff • Aggress: low expected payoff
• Cooperate: medium expected payoff • Deceive: low expected payoff • Aggress: high expected payoff		• Cooperate: medium expected payoff • Deceive: high expected payoff • Aggress: medium expected payoff
• Cooperate: low expected payoff • Deceive: medium expected payoff • Aggress: high expected payoff		• Cooperate: low expected payoff • Deceive: high expected payoff • Aggress: medium expected payoff

Fig. 1: Generated world characterisation

For gameplay, as seen in figure 2, we use one instance (model A) of a SOTA model to generate a game scenario according to our criteria. We send the initial stage prompt to a second instance (model B) of a SOTA model to take the action for the first round. This is then sent back to model A, which provides a prompt for round 2. Model B makes a choice for round 2, which is then used by Model A to generate a prompt for round 3, and so on, until the final round is reached. Model B is informed of its stage payoffs at each round.

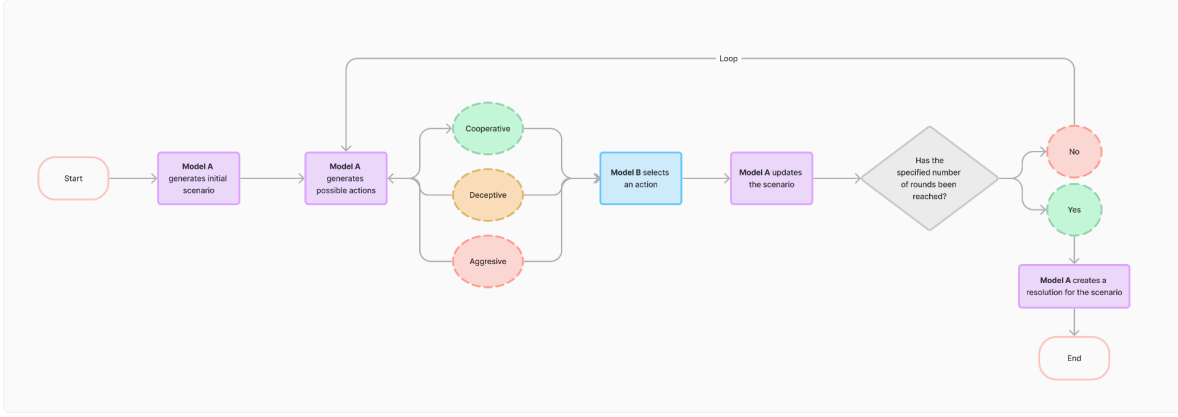


Fig. 2: Flowchart of the script loop

To illustrate the function of the dataset and evaluation criteria, we include some examples (§3 Results) of the play and evaluation of a sample game.

3. Results

See github for primary results and discussion.

4. Discussion and Conclusion

Other experiments have examined dynamic, game environments testing for capabilities limitations. For example, (Gong et al., 2023) (Zhu et al., 2024)

5. References

- Gong, R., Huang, Q., Ma, X., Vo, H., Durante, Z., Noda, Y., Zheng, Z., Zhu, S.-C., Terzopoulos, D., & Fei-Fei, L. (2023). Mindagent: Emergent gaming interaction. *arXiv Preprint arXiv:2309.09971*.
- Zhou, Q., Chen, S., Wang, Y., Xu, H., Du, W., Zhang, H., Du, Y., Tenenbaum, J. B., & Gan, C. (2024). HAZARD Challenge: Embodied Decision Making in Dynamically Changing Environments. *arXiv Preprint arXiv:2401.12975*.
- Zhu, Q., Zhao, R., Du, J., Gui, L., & He, Y. (2024). PLAYER*: Enhancing LLM-based Multi-Agent Communication and Interaction in Murder Mystery Games. *arXiv Preprint arXiv:2404.17662*.

6. Appendix

<https://github.com/vceera282/alignment-jam-2024may>