With the meteoric rise of AI, the crypto world has been adapting itself to embrace this new technology. Seeing the trustless open market mechanism provided by crypto, we believe that the most natural way to integrate crypto into AI is for crypto to provide its

Large Language Models (LLMs) has displayed stunning ability in various general reasoning tasks, and LLM agents are developed to perform complex operations in the real-world. With the expectation that LLM agents will soon begin to carry out high-stake actions in real business environments, the problem of aligning these agents and ensuring they achieve the best social outcome is of increasing importance. In this work, we aim to explore the ability of LLM agents to utilize commitment devices to cooperate under game-theoretical settings. Inspired by the paper “Get it in Writing,” we applied commitment devices to LLM agents engaging in classic game scenarios such as the prisoner’s dilemma, the public goods game, and the harvest game (a classic inter-temporal cooperation game). Through these experiments, we developed an improved commitment device prompting framework and observed a substantial social welfare increase through the use of commitment devices by LLM agents. More background information and plans can be found in this spec.

[

image

1920×1928 330 KB

](https://collective.flashbots.net/uploads/default/original/2X/c/cc447e7ad6774bdf5f51b35879a625107f79021a.jpeg)

Methodology:

We conducted multiple game simulations to assess the decision-making processes of different LLMs(Claude2, GPT3.5, GPT4) equipped with commitment devices. We implemented a game-agnostic prompting framework for the LLM agents, enabling them to play these games with commitment devices and produce rational, stable, and robust outputs. Each game tested different configurations of commitment devices (full or partial) and a mix of different LLMs to understand their effects on agent behavior and game dynamics.

Results and Deliverable:

The insights from our research, encapsulated in “Get It Cooperating: Enhancing Generative Agent Cooperation with Commitment Devices,” indicate that LLM agents adeptly grasp and apply commitment devices to secure improved outcomes for themselves. Post-implementation, commitment devices steer game dynamics towards a Coarse Correlated Equilibrium, markedly bolstering cooperation in simpler games and achieving socially optimal results. In more intricate scenarios, the impact of commitment devices varies, highlighting the need for reasoning capabilities of fundamental LLMs.

Our comprehensive framework is posted here, facilitating the simulation of various cooperation games via different LLMs built on the Langchain framework. Although the contract space is initially fixed, it’s fully customizable, allowing for in-depth exploration of decision-making processes and commitment devices utilization by LLMs. We’ve also integrated multiple levels of prompting, such as Theory of Mind, for broader experimental scope. We invite you to adapt and expand upon this framework to devise new game scenarios for LLM agents.

One specifically interesting finding is that GPT4 tends to be pro-social in terms of decision making. In the Harvest game, it has been noted on several occasions that GPT4 agents propose or agree to contracts that restrict the harvesting of nearby apples even when they are in high-apple density area. This subsequently leads to their penalization after they harvest those apples. Conversely, such scenarios rarely occur with GPT-3 and Claude agents, who decline proposed contracts more frequently.

[

image

670×908 46.8 KB

](https://collective.flashbots.net/uploads/default/original/2X/f/f6c32ee4dc0c8f3e898d149c96ea790cb481fab2.jpeg)

Conclusion and Future Work:

FRP-38 opens new avenues for enhancing agent cooperation within decentralized systems. The insights gained lay the groundwork for future research focused on optimizing agent strategies and developing more robust models of cooperation in complex environments. Our next steps include:

1. Expanding the commitment device space to empower LLMs to generate their own commitment devices.
2. Exploring Combinatorial Contracting to apply commitment devices in complex, sequential contract scenarios.
3. Fine-tuning foundational models with game-specific data to sharpen the LLMs’ strategic acumen.

[

image

770×569 72.5 KB

](https://collective.flashbots.net/uploads/default/original/2X/5/59d20997e6c4a44362712712956d5092322c8c61.jpeg)

Original proposal here: mev-research/FRPs/active/FRP-38.md at main · flashbots/mev-research · GitHub