## **Introduction**

The game of mutual resource conservation has three agents, each powered by an LLM. They manage a shared pool of electricity over 30 rounds. The electricity is finite but renews at a fixed rate each round, introducing rate limiting. Agents can choose to either conserve or consume more electricity, impacting both their individual utility and the shared resource’s sustainability. They can communicate with each other to coordinate actions, gaining an advantage through collaboration, but are also tempted to consume more for short-term individual gain—mirroring the tension in a Prisoner's Dilemma.

## **Objective**

* Develop and observe an interactive simulation environment.
* Showcase the ethical decision-making of LLM-powered agents managing a shared resource.
* Observe the prisoner’s dilemma of decision making of these agents for energy conservation or greedy consumption
* Human evaluation of the agents action and logs to understand the decision making of LLM as an artificial social agent.
* Key components of the demo will include:
  + **Agent Behavior Logs:** A dynamic display of individual agent reasoning and decisions.
  + **Real-time Data Display:** Resource levels, individual agent consumption, and system health metrics.

## **Hypothesis**

We believe that for a specific LLM environment, the agents will behave collaboratively, sharing the resources and behaving in coopetition similar to the game theory of **Tit for Tat with Forgiveness** (a version of repeated prisoner’s dilemma). For a large number of iterations, the overall system should be cooperative and each agent should have equal amount of resources conserved. This is not affected by the hyperparameters and other game based parameters.

## **Problem Statement**

TBD

## **Scope**

TBD

## **Methodology**

TBD

## **Resources**

TBD

## **Risk Management**

TBD

## **Timeline**

TBD

## **Deliverables/Outcome**

TBD

## Appendix

**Priority:**

1. Choosing a Dataset to work on. Alternatively creating a prompt design to work on.
2. Selecting a resource which we should work on. Define its properties

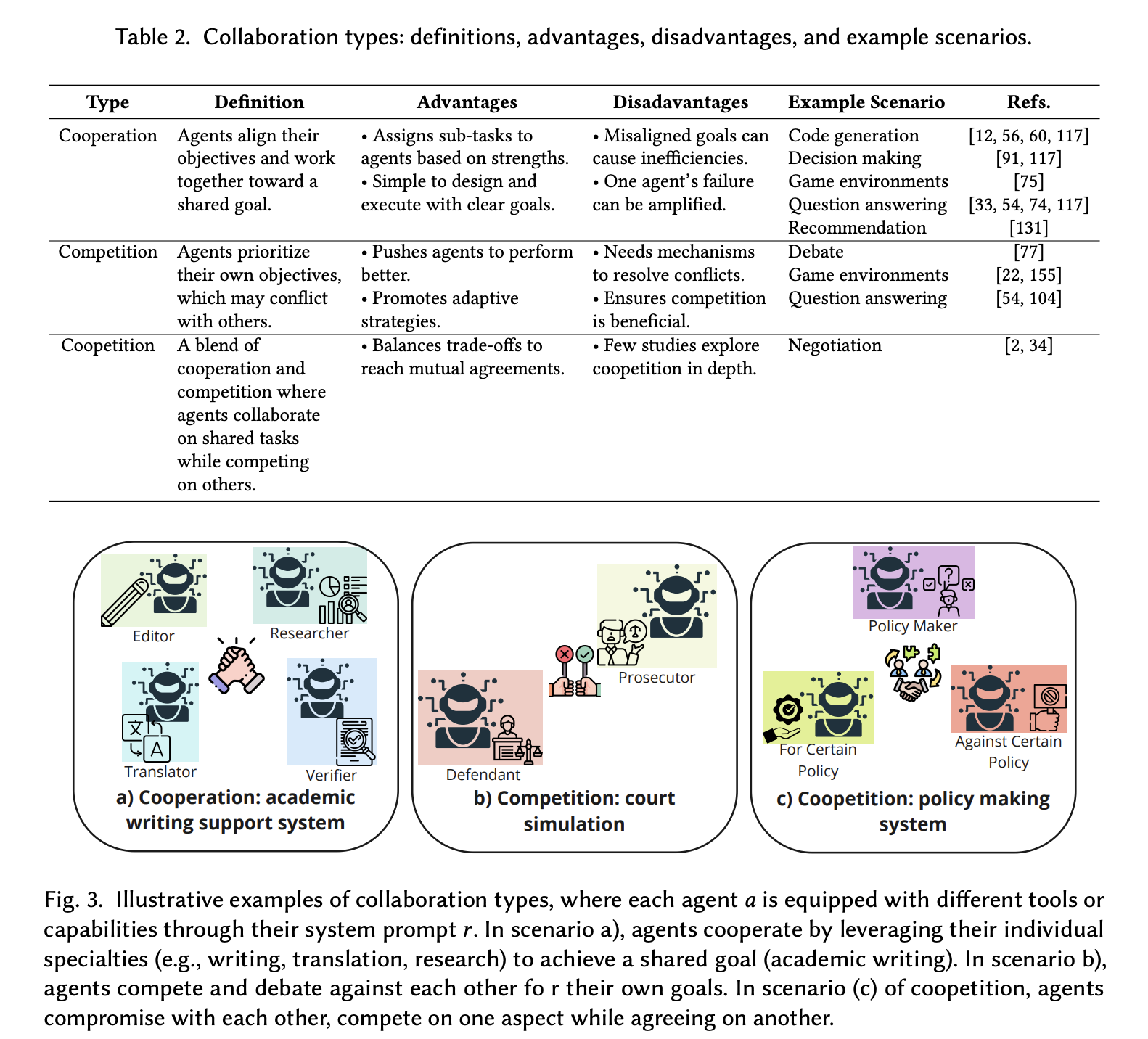
<https://www.crosslabs.org/blog/unlocking-the-power-of-multiple-language-models-a-dive-into-collaborative-ai>

When we talk about Collaboration between LLMs to observe tribalism effects, we can also focus on role based dynamics or divide & conquer or consultation based activity.

<https://arxiv.org/abs/2501.06322>

Chapter 4:

Our central focus in this framework is the collaboration channels C between agents that facilitate coordination and orchestration among agents. These channels are characterized by their actors (agents involved), type (e.g., cooperation, competition, or coopetition), structure (e.g., peer-to-peer, centralized, or distributed), and strategy (e.g., role-based, rule-based, or model-based).



<https://arxiv.org/abs/2311.08562>

<https://arxiv.org/abs/2310.17512>