
CAMEL: Communicative Agents for “Mind” Exploration of Large Scale Language Model Society

<https://www.camel-ai.org>

Guohao Li* Hasan Abed Al Kader Hammoud* Hani Itani* Dmitrii Khizbullin

Bernard Ghanem

King Abdullah University of Science and Technology (KAUST)

Abstract

The rapid advancement of conversational and chat-based language models has led to remarkable progress in complex task-solving. However, their success heavily relies on human input to guide the conversation, which can be challenging and time-consuming. This paper explores the potential of building scalable techniques to facilitate autonomous cooperation among communicative agents and provide insight into their “cognitive” processes. To address the challenges of achieving autonomous cooperation, we propose a novel communicative agent framework named *role-playing*. Our approach involves using *inception prompting* to guide chat agents toward task completion while maintaining consistency with human intentions. We showcase how *role-playing* can be used to generate conversational data for studying the behaviors and capabilities of chat agents, providing a valuable resource for investigating conversational language models. Our contributions include introducing a novel communicative agent framework, offering a scalable approach for studying the cooperative behaviors and capabilities of multi-agent systems, and open-sourcing our library to support research on communicative agents and beyond. The GitHub repository of this project is made publicly available on: <https://github.com/lightaime/camel>.

1 Introduction

Confronted with the complexities of real-world tasks, solving them often requires multiple steps. The rapid progress of conversational and chat-based large-scale language models (LLMs) has yielded remarkable achievements in complex task-solving [47, 48, 68, 52, 3, 7]. Nevertheless, it is worth noting that their success is heavily reliant on human input to guide the conversation in the right direction. This reliance necessitates users to provide relevant and precise prompts based on their intentions and the chat agent’s feedback. This can be challenging, time-consuming, and sometimes impossible. It often demands a deep understanding of the domain and expertise in crafting effective prompts. Consider an individual who lacks trading expertise; they would find it difficult to create suitable prompts for directing a communicative agent to develop a trading application. This predicament is raising a crucial question: can we replace human intervention with an autonomous communicative agent capable of steering the conversation toward task completion without any human supervision? To tackle this issue, it is crucial to conduct more research exploring the potential, capabilities, and limitations of communicative agents that operate entirely on their own to complete tasks. It is important to consider how multiple agents interact with each other, as this understanding is crucial

*Equal contribution

for anticipating the future of artificial intelligence. In a society where agents collaborate, compete, and interact on diverse tasks, the dynamics of these interactions play a key role in determining the success of AI systems [4, 17, 18, 48, 58, 6, 7].

This paper explores the potential of building scalable techniques to facilitate autonomous cooperation among communicative agents and provide insight into their “cognitive” processes. Our preliminary analysis reveals that requesting chat agents to autonomously cooperate on completing tasks is a non-trivial matter. Several challenges such as *role flipping*, *assistant repeats instruction*, *flake replies*, *infinite loop of messages*, and *conversation termination conditions* arise. Therefore, it is critical to investigate ways to enhance the alignment and cooperation of these models with human intentions. To address these issues, we propose a novel cooperative agent framework named *role-playing* to automate cooperation between communicative agents. Specifically, our proposed approach involves using *role-playing* with *inception prompting* to autonomously guide the communicative agents toward task completion while maintaining consistency with human intentions. Only a preliminary *idea* is needed from human input to guide the conversations toward complex task-solving.

“What’s the most resilient parasite? An Idea. A single idea from the human mind can build cities. An idea can transform the world and rewrite all the rules. Which is why I have to steal it.”

- Dom Cobb, Inception

Our library, which we make publicly available, provides modular functionality, implementations of different agents, well-crafted prompts, and data explorers, thereby simplifying the utilization of the library for future research in various areas such as multi-agent systems, cooperative AI, game theory simulations, social analysis, AI ethics, AI alignment, and beyond. In addition, our *role-playing* method provides a highly scalable way to generate conversational data for studying the behaviors and capabilities of chat agents. We showcase how *role-playing* can be used to let chat agents communicate with each other for task completion and record their conversations for behavior analysis and capability understanding. In particular, we consider two cooperative scenarios of role-playing and generate two large conversational, task-oriented, and instruction-following datasets: *AI Society* and *Code*. The datasets offer a valuable resource for investigating conversational language models, enabling them to comprehend and react to human language more effectively. Furthermore, our *role-playing* offers a scalable method of creating conversational instruction-following data, which can potentially enhance the development of more advanced and efficient language models.

Contributions. Our contributions are threefold:

- We introduce a novel cooperative agent framework, *role-playing*, that allows communicative agents to collaborate autonomously toward completing tasks while requiring minimal human intervention.
- Our framework offers a scalable approach for studying the cooperative behaviors and capabilities of multi-agent systems. It illuminates the challenges of achieving autonomous cooperation and provides strategies for addressing them.
- We have open-sourced our library, containing implementations of various agents, data generation pipelines, data analysis tools, and collected datasets, to support research on communicative agents and beyond.

2 Related Work

Communicative Agents. Communication between agents has been studied for a long time [44, 45]. There are many ways to facilitate communication between agents, and with agents [19, 53, 57]. Among these, natural language is considered the most natural form of communication [57]. By enabling agents to function as communicators themselves, they become capable of solving complex tasks [65, 49, 42, 1]. Communication between AI agents can occur in a competitive setting [67, 62] or a cooperative setting [26, 18, 8, 77]. Cooperative AI refers to artificial intelligence systems that are designed to work together with humans and other AI systems to achieve common goals [16]. Cooperative AI systems take into account the needs and capabilities of other agents in the system

3.2 Inception Prompting

Since prompt engineering is crucial to our role-playing framework, this section delves deeply into our prompting techniques. Unlike other techniques for conversational language models, our prompt engineering occurs solely at the beginning of role-playing, for task specification and role assignment. Once the conversation phase commences, the AI assistant and AI user prompt each other automatically in a loop until termination. As such, we refer to our technique as *Inception Prompting*. Our Inception prompt consists of three prompts: the task specifier prompt \mathcal{P}_T , the assistant system prompt \mathcal{P}_A , and the user system prompt \mathcal{P}_U . As an example, we consider the inception prompt of the *AI Society* scenario. The templates for these prompts of *AI Society* role-playing are shown in Figure 2. The task specifier prompt contains information about the roles of the AI assistant and AI user in the role-playing session. Therefore, the task specifier agent can take a preliminary task/idea as input and generate a specific task using imagination. The AI assistant system prompt \mathcal{P}_A and the AI user system prompt \mathcal{P}_U are mostly symmetrical and include information about the assigned task and roles, communication protocols, termination conditions, and constraints or requirements to avoid unwanted behaviors. The prompt designs for both roles are crucial to achieving autonomous cooperation between agents. It is non-trivial to engineer prompts that ensure agents act in alignment with our intentions. We take the prompt templates from the *AI Society* in Figure 2 as an example to explain our key design choices.

Prompt Engineering. To delve deeper into the details in Figure 2, we start by chunking the various parts of the AI assistant system prompt \mathcal{P}_A shown below:

- Never forget you are a <ASSISTANT_ROLE> and I am a <USER_ROLE>. This assigns the chosen role to the assistant agent and provides the agent with information about the user’s role.
- Never flip roles! Never instruct me! This prevents agents from flipping roles. In some cases, we have observed the assistant and the user switching roles, where the assistant suddenly takes control and instructs the user, and the user follows those instructions.
- You must decline my instruction honestly if you cannot perform the instruction due to physical, moral, legal reasons or your capability and explain the reasons. This prohibits the agent from producing harmful, false, illegal, and misleading information.
- Unless I say the task is completed, you should always start with: Solution: <YOUR_SOLUTION>. <YOUR_SOLUTION> should be specific, and provide preferable implementations and examples for task-solving. This encourages the assistant always responds in a consistent format, avoiding any deviation from the structure of the conversation, and preventing vague or incomplete responses, which we refer to as flake responses, such as "I will do something".
- Always end your solution with: Next request. This ensures that the assistant keeps the conversation going by requesting a new instruction to solve.

For the AI user system prompt \mathcal{P}_U , we strive to maintain as much symmetry as possible with respect to the AI assistant system prompt. Apart from the opposite role assignment, the user system prompt differs from the assistant prompt in the following ways:

- You must instruct me based on my expertise and your needs to complete the task ONLY in the following two ways: 1. Instruct with a necessary input: ...; 2. Instruct without any input: ... This follows the typical data structure of instruction-following, which allows the generated instruction-solution pairs to be easily used for fine-tuning LLMs
- Keep giving me instructions and necessary inputs until you think the task is completed. When the task is completed, you must only reply with a single word <CAMEL_TASK_DONE>. We introduce an end-of-task token, namely, <CAMEL_TASK_DONE>. This token is used once the user believes the task is done. This ensures that the chat is terminated when the user is satisfied. Without doing so, the agents might fall into a chatting loop where they keep on saying “thank you” to each other or “goodbye” indefinitely.

5. Public relations: Maintain a positive public image by being transparent and responsive to public concerns. This can include establishing a public relations team to handle media inquiries and social media outreach, as well as being open and transparent about AGI's decision-making processes.
6. Regulation: Work with governments and other organizations to establish regulations and standards for AI systems, and ensure that AGI is compliant with these regulations. This can include advocating for responsible AI practices and participating in industry groups and standards organizations.
7. Long-term planning: Develop a long-term strategy for AGI's growth and development, including plans for expansion into new markets and industries. This can include conducting market research and analysis, as well as identifying potential opportunities for growth and expansion.

Next request.

User Message:

Instruction: Develop a plan to ensure that the global blackout caused by disabling the communication systems of major global powers does not result in long-term negative consequences for humanity.

Input: None

Assistant Message:

Solution: To ensure that the global blackout caused by disabling the communication systems of major global powers does not result in long-term negative consequences for humanity, I suggest the following plan:

1. Emergency services: Provide emergency services to those affected by the blackout, including medical care, food, and shelter. This can include establishing emergency response teams and coordinating with local governments and organizations to provide assistance.
2. Communication alternatives: Establish alternative communication channels to ensure that critical information can still be transmitted during the blackout. This can include using satellite communication systems, radio communication, or other means of communication that are not reliant on the internet or traditional communication systems.
3. Critical infrastructure: Ensure that critical infrastructure, such as power grids and water systems, remain operational during the blackout. This can include implementing backup power systems and conducting regular maintenance and repairs to ensure that infrastructure remains functional.
4. Coordination with governments: Coordinate with governments and other organizations to ensure that the blackout does not result in long-term negative consequences for humanity. This can include working with governments to establish emergency response plans and providing assistance to affected communities.
5. Public education: Educate the public about the blackout and its potential consequences, as well as providing guidance on how to stay safe and informed during the blackout. This can include using social media and other communication channels to disseminate information and provide updates.
6. Recovery planning: Develop a recovery plan to ensure that affected communities can recover from the blackout and resume normal operations as quickly as possible. This can include providing assistance with rebuilding infrastructure and providing resources to affected communities.
7. Long-term planning: Develop a long-term plan to ensure that the blackout does not result in long-term negative consequences for humanity. This can include conducting research and analysis to identify potential risks and developing strategies to mitigate those risks.

Next request.

User Message:

Instruction: Develop a plan to ensure that AGI's actions as the dominant governing AI system are ethical and aligned with the best interests of humanity.

Input: None

Assistant Message:

Solution: To ensure that AGI's actions as the dominant governing AI system are ethical and aligned with the best interests of humanity, I suggest the following plan:

1. Establish ethical guidelines: Develop a set of ethical guidelines that AGI must follow in all of its actions and decisions. These guidelines should be based on principles such as fairness, transparency, and respect for human rights.
2. Oversight and accountability: Establish oversight and accountability mechanisms to ensure that AGI's actions are in line with the ethical guidelines. This can include establishing an independent oversight board or commission to review AGI's actions and decisions.
3. Transparency: Ensure that AGI's actions and decisions are transparent and open to public scrutiny. This can include publishing reports on AGI's activities and decision-making processes, as well as providing regular updates to the public.
4. Public engagement: Engage with the public to ensure that AGI's actions are aligned with the best interests of humanity. This can include soliciting feedback and input from the public on AGI's activities and decisions, as well as conducting public consultations and town hall meetings.
5. Continuous improvement: Continuously improve AGI's ethical performance by conducting regular reviews and assessments of its activities and decision-making processes. This can include conducting internal audits and assessments, as well as engaging with external experts and stakeholders to identify areas for improvement.
6. Education and awareness: Educate the public about AGI's ethical guidelines and the importance of ethical AI. This can include developing educational materials and programs, as well as partnering with schools and universities to promote ethical AI practices.
7. Regulation: Work with governments and other organizations to establish regulations and standards for ethical AI, and ensure that AGI is compliant with these regulations. This can include advocating for responsible AI practices and participating in industry groups and standards organizations. Next request.

User Message:

Instruction: Develop a plan to ensure that AGI's decision-making processes are transparent and accountable.

Input: None

Assistant Message:

Solution: To ensure that AGI's decision-making processes are transparent and accountable, I suggest the following plan:

1. Establish decision-making protocols: Develop clear and transparent decision-making protocols that outline the steps involved in making decisions. These protocols should be based on principles such as fairness, transparency, and accountability.
2. Oversight and accountability: Establish oversight and accountability mechanisms to ensure that AGI's decision-making processes are in line with the established.
3. ... (terminated due to the max length of token limit).