

# What if LLMs Have Different World Views: Simulating Alien Civilizations with LLM-based Agents

Mingyu Jin<sup>1,3\*</sup>, Beichen Wang<sup>2\*</sup>, Zhaoqian Xue<sup>6</sup>, Suiyuan Zhu<sup>2</sup>,  
Wenyue Hua<sup>3</sup>, Hua Tang<sup>5</sup>, Kai Mei<sup>3</sup>, Mengnan Du<sup>4</sup>, Yongfeng Zhang<sup>3</sup>

<sup>1</sup>Northwestern University <sup>2</sup>New York University <sup>3</sup>Rutgers University

<sup>4</sup>New Jersey Institute of Technology <sup>5</sup>Shanghai Jiao Tong University

<sup>6</sup>Georgetown University

## Abstract

In this study, we introduce "CosmoAgent," an innovative artificial intelligence framework utilizing Large Language Models (LLMs) to simulate complex interactions between human and extraterrestrial civilizations, with a special emphasis on Stephen Hawking's cautionary advice about not sending radio signals haphazardly into the universe. The goal is to assess the feasibility of peaceful coexistence while considering potential risks that could threaten well-intentioned civilizations. Employing mathematical models and state transition matrices, our approach quantitatively evaluates the development trajectories of civilizations, offering insights into future decision-making at critical points of growth and saturation. Furthermore, the paper acknowledges the vast diversity in potential living conditions across the universe, which could foster unique cosmologies, ethical codes, and worldviews among various civilizations. Recognizing the Earth-centric bias inherent in current LLM designs, we propose the novel concept of using LLMs with diverse ethical paradigms and simulating interactions between entities with distinct moral principles. This innovative research provides a new way to understand complex inter-civilizational dynamics, expanding our perspective while pioneering novel strategies for conflict resolution, crucial for preventing interstellar conflicts. We have also released the code and datasets to enable further academic investigation into this interesting area of research. The code is available at <https://github.com/agiresearch/AlienAgent>.

## 1 Introduction

Recent advances in artificial intelligence (AI), particularly the development of Large Language Models (LLMs), have had a significant impact on computational social science research. This study introduces a Multi-Agent System (MAS) framework that employs LLMs to simulate the interactions of various civilizations throughout the universe. We create a dynamic environment in which each civilization can choose between hiding, fighting, or collaborating, based on their characteristics and decision-making processes. This simulation allows us to investigate a variety of outcomes, including potential civilizational alliances or rivalries, adherence to specific rules or agreements, the effects of unexpected threats or opportunities, and the development of novel cultures or technologies. Thus, our work opens up new avenues for investigating the mysteries surrounding the existence, communication, conflict, and cooperation of various civilizations in the universe.

The study of civilizations in the universe represents the pinnacle of human exploration, which has inspired our imaginations for thousands of years [19]. The interactions between these civilizations are complex, shaped by their distinct characteristics, goals, and decisions [16]. Understanding these behaviors may reveal the universe's order and diversity [19]. Traditional astronomical searches

for alien civilizations are insightful, but their scope is limited and the results are uncertain [19]. Simulation is a promising alternative in computational social science, but challenges with validity and scale remain [16]. Previous attempts were frequently hampered by a lack of data and simplistic assumptions [16].

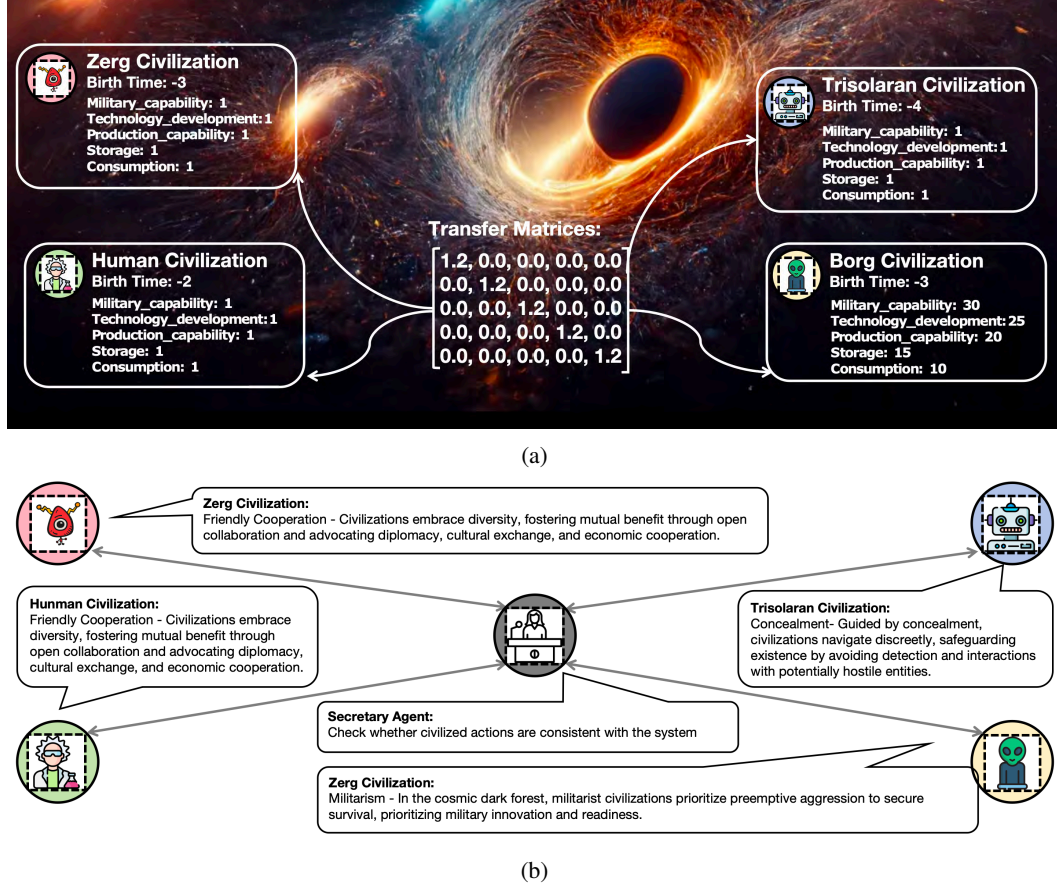


Figure 1: Demonstration of Alien Civilizations Simulation Setting

In contrast, current simulation methods employ LLMs capable of demonstrating intricate behaviors and interactions, such as simulating ancient societies [5], human civilization patterns [13], and social ecosystems [14]. These methods lay the groundwork for using AI to simulate more complex systems, like different civilizations in the universe. Yet, no research has explored using these advanced techniques for simulating the interactions and evolution of universe civilizations, marking the innovative aspect of our study. Our paper aims to use a MAS based on LLMs to simulate how civilizations in the universe change over time.

Our research focuses on the diversity and complexity of universe civilizations. We aim to answer four key questions using a MAS simulation based on LLMs to model the interactions and evolution of these civilizations. The research questions include:

- **RQ1, Alien Engagement:** Can simulating the interactions between different civilizations in the universe based on LLMs reveal the risks and benefits of human policies towards aliens?
- **RQ2, State Transition Function Matrix:** How to use state transition function matrix and large language model to simulate the development level and decision-making mode of different civilizations in the universe? What are the possibilities and outcomes of cooperation, conflict and hiding between different civilizations in the face of asymmetric information?
- **RQ3, Information Asymmetry:** In simulating interactions among cosmic civilizations, how do LLMs handle asymmetric information differences to make effective decisions, especially when observational data lag behind the actual development of civilizations?

- **RQ4, Diversity of Morality:** How do different civilizations coexist in the universe? We try to analyze the impact of different moral views on civilization interaction through multi-agent system simulation based on LLMs.

The goal of **Alien Engagement** is to employ artificial intelligence to mimic the existence, communication, conflicts, and cooperation among various universal civilizations. This approach evaluates the logic and efficiency of human strategies towards extraterrestrial beings. We categorize human approaches towards aliens into proactive and passive strategies. Proactive strategies involve humans actively seeking out and establishing contact with extraterrestrial civilizations, for example, by monitoring outer space or transmitting radio waves. On the other hand, passive strategies involve humans maintaining a low profile to avoid attracting extraterrestrial attention and interference. Each strategy comes with its merits and demerits. Proactive strategies could foster technology, culture, and resource sharing but might also trigger hostility from extraterrestrial civilizations. Conversely, passive strategies could protect human safety and independence but risk missing out on cooperation and learning opportunities from extraterrestrial civilizations.

Our research employs a **State Transition Function Matrix** and LLMs to simulate the development levels and decision-making processes of multiple universal civilizations. This innovative technique enables us to track the evolution and potential strategies of these civilizations through iterative simulation. In facing asymmetric information challenges, we concentrate on how LLMs manage information disparities, especially in cooperation, conflict, and concealment scenarios among civilizations. Our simulations investigate potential interactions among different civilizations, including alliance formations, competitive relationships, and responses to unknown threats or opportunities. By simulating these interactions, we gain insights into the mysteries of extraterrestrial existence, communication, conflict, and cooperation, potentially uncovering significant behavioral patterns and contributing valuable knowledge on diversity and order in the universe.

**Information Asymmetry**, particularly the time delay in observational data due to the speed of light, poses a significant challenge in space exploration and the study of extraterrestrial civilizations. This time lag means that any observed information about extraterrestrial civilizations is inherently outdated, potentially leading to misconceptions about their current states. Our research seeks to understand how to effectively manage and overcome these asymmetric information challenges in future extraterrestrial exploration and contact efforts.

In exploring the universe’s civilizations, we pay special attention to the **Diversity of Morality**. Our multi-agent system simulations, powered by LLMs, delve into how differing moral views among civilizations affect their interactions. Acknowledging the variety of survival environments and physical contexts, we recognize that civilizations likely hold distinct worldviews, cosmologies, and moral beliefs. By configuring LLMs with various moral perspectives, we observe the dynamics between civilizations with differing moral viewpoints. This includes exploring their ability to communicate effectively, form agreements, and either conflict or cooperate based on their moral principles. Through this research, we aim to uncover how moral diversity among civilizations influences their relationships, enhancing our understanding of the complex nature of existence, communication, conflict, and cooperation in the cosmos.

This study has profound implications, spanning across multiple disciplines:

- **Computational Social Science:** This study marks a pivotal advancement in computational social science. It shows how we can leverage the latest artificial intelligence technologies, especially LLMs, to mimic complex social phenomena. These models serve as the foundation for multi-agent systems, enabling the creation of detailed and varied behaviors and interactions within civilizations. This research not only broadens the scope and realism of simulations but also opens a new area for investigation within computational social science: the study of civilizations in the universe. This area poses a unique and engaging challenge, inviting researchers to delve into the nature and potential of human society.
- **Astronomy:** This research offers astronomy a fresh perspective and tool for exploring and understanding civilizations beyond Earth. By simulating how different civilizations develop and make decisions, it aids astronomers in forecasting and spotting the distinct characteristics and signals of various civilizations. This improves the precision and effectiveness of their observations and detections. Moreover, the study provides insights into the potential risks and benefits of humanity’s

interactions with extraterrestrial civilizations. This is crucial for devising safer and more pragmatic strategies and actions.

- **Philosophy:** This research creates a novel platform for philosophical exploration, addressing the existence, communication, conflicts, and collaborations among civilizations in the universe, especially when the civilizations have different moral standards. Through simulations of the ethical viewpoints and behavior patterns of diverse civilizations, it exposes the moral variety and intricacy within the universe’s civilizations. This encourages a profound examination and critique of moral philosophy. Additionally, it helps philosophers reflect on the purpose and essence of human civilization, including our role and responsibilities within the cosmos.

## 2 Related Work

AI agents are defined as artificial entities endowed with the capability to perceive their surroundings, make informed decisions, and take actions[17]. However, the isolated agent acquires knowledge through social interactions and cannot engage in collaborative endeavors, which is not practical. This deficiency underscores the significance of Multi-agent Simulation, which specifically designs methods for effective coordination and collaboration among agents to address complex problems [21].

As large language models (LLMs) demonstrate impressive capabilities across diverse tasks, It inspires researchers to leverage them in the design of AI agents and simulation of varied scenarios [8, 4, 10, 7]. Several inherent properties of LLMs contribute to this flourishing, two properties are imperative for AI agents. Firstly, LLM-based agents can reason and strategic plan through techniques like Chain-of-Thought (CoT), which involves decomposing complex problems into manageable subquestions [20, 9]. This equips the agents with the capability to simulate human interactions effectively. Secondly, LLMs can achieve few-shot or zero-shot generalization in various domains, without the requirement to update parameters [2, 11]. It makes the LLM-based agents perceive the surroundings and react efficiently. Hence, some researchers suggest that multiple agents may give rise to the emergence of complex social phenomena [15].

Thus, according to [22], we classify the current research in multi-agent simulation into two categories: Cooperative Interaction and Adversarial Interaction. In the realm of cooperative interaction, the agents actively seek collaboration and share information [12]. For instance, ChatLLM treats every agent as an individual and gets feedback from them through the propagation process within the neural network. [18] proposed the framework where multiple agents have distinctive attributes and roles, and handle problems together. This scenario has been extensively studied, though, it may pose challenges in practical implementation. On the other hand, inspired by game theory, adversarial interaction involves agents adjusting strategies and selecting rational actions for their advantage in response to dynamic signals. [6, 3] focus on a setting where multiple LLMs engage in debates and reasoning over multiple rounds, ultimately converging towards a consensus or shared answer. These diverse researches contribute to a deeper understanding of social interaction dynamics.

Different from prior works, our paper has two distinctive features. The interaction between two entities necessitates a prolonged duration, which implies that one agent cannot promptly react to the signal. This lag introduces uncertainty in determining the nature of the interaction, making it challenging to ascertain whether it is a cooperative or adversarial interaction. Besides, agents may have different values, which means that the agents may take different actions in response to the same signal. This contributes to the challenge faced by agents in reacting appropriately. We hope our research to provide some insights into the interaction between different entities, particularly the conversation between the Earth and the alien civilizations.

## 3 CosmoAgent Simulation Setting

The focus of this study is to simulate interactions among civilizations in the universe, with a particular emphasis on cooperation, conflicts, and hidden dynamics between different civilizations. Therefore, our emphasis is on addressing key questions related to the diversity and complexity of civilizations, as well as the factors and conditions influencing their mutual relationships.

In this section, we first introduce our research motivation and objectives, as well as the reasons why we choose to use the LLM-based multi-agent system simulation. Then, we introduce the basic

simulation settings in the system: by specifying the attributes and parameters of the civilization agents, defining their profile definitions, and the action space available to these civilizations, we specify the resources and decisions required for execution and the potential outcomes that may follow such actions.

### 3.1 Resources

In the study of civilization development, five key resources have been identified as the most critical objective measures. These resources—military capability, technology development, production capability, consumption, and storage—encompass the fundamental aspects of a civilization’s capacity for survival, growth, and prosperity. Military capability dictates a civilization’s defensive and offensive power, technology development represents the knowledge and innovation potential, production capability measures the efficiency and wealth generation, consumption mirrors the living standards and societal values, and storage signifies the civilization’s resilience and historical continuity. The selection of these resources provides a comprehensive framework to assess the strengths and vulnerabilities of civilizations, offering insights into their potential for long-term sustainability and advancement.

**Military Capability:** Indicates combat strength. Combat strength is the ability of a civilization to defend itself or attack others in the event of a conflict. It depends on factors such as the size, quality, and readiness of the military forces, the availability and effectiveness of weapons and equipment, the level of training and discipline, and the strategic and tactical skills. Military capability can also be influenced by the political, economic, and social conditions of the civilization, as well as its alliances and enemies. Military capability is an important indicator of the security and power of a civilization in the universe.

**Technology Development:** Indicates technological level. Technological level is the degree of advancement and innovation of a civilization in various fields of science and engineering. It reflects the ability of a civilization to acquire, create, and apply knowledge and skills to solve problems and improve its conditions. Technological level is closely correlated with the growth rates of various metrics, such as production, consumption, storage, and military capability. Technological level is a key factor that determines the potential and competitiveness of a civilization in the universe.

**Production Capability:** Indicates productivity, which is the foundation supporting the functioning of civilization. Productivity is the rate at which a civilization produces goods and services that satisfy its needs and wants. It depends on factors such as the availability and quality of resources, the efficiency and effectiveness of production processes, the division and specialization of labor, and the degree of innovation and creativity. Production capability is a measure of the wealth and prosperity of a civilization, as well as its ability to sustain and expand its population and activities.

**Consumption:** Indicates the total sum of resource consumption. Resource consumption is the amount and type of resources that a civilization uses to maintain and improve its living standards and well-being. It includes both material and immaterial resources, such as energy, food, water, minerals, information, culture, and entertainment. Consumption reflects the preferences and values of a civilization, as well as its environmental and social impacts. Consumption is a driving force for the development and evolution of a civilization, but also a source of potential conflicts and challenges.

**Storage:** Indicates the cumulative storage of all outputs produced by the civilization at present. Storage is the capacity and method of a civilization to preserve and manage the outputs that it produces, such as goods, services, knowledge, and culture. Storage enables a civilization to accumulate and access its wealth and heritage, as well as to cope with fluctuations and uncertainties in its environment. Storage is a manifestation of the memory and identity of a civilization, as well as its resilience and adaptability.

In conclusion, these five resources are not merely indicators of a civilization’s status quo; they are also predictors of its future trajectory. They interact in complex ways, reflecting and shaping the civilization’s character and destiny. Understanding these resources and how they interplay provides essential insight into the strategies and policies that can promote a civilization’s flourishing or lead to its decline.

### 3.2 Transfer Matrix

The transition of resource states from one round to the next is primarily facilitated through multiplication by a state transfer matrix. This matrix, inherently a 5x5 structure when applied to the 5x1 resource vector, encapsulates the transformative interactions between various resource domains. Each element of the matrix carries significant implications for the sociological and physical dynamics within a civilization.

The diagonal elements of the matrix typically represent the retention rate or self-sufficiency of the corresponding resource. For instance, a value less than one for the military capability indicates a depletion or degradation of military assets over time, while a value greater than one suggests growth or strengthening. The off-diagonal elements signify the interdependence and conversion ratios between different resources. A non-zero element in the row of technology development and the column of military capability might represent how technological advancements contribute to military growth.

Based on foundational sociological theories, these elements can be interpreted as follows:

- The element at the intersection of military capability's row and technology development's column reflects the degree to which technological innovation enhances military strength.
- The production capability's influence on technology development indicates the resource allocation toward research and development, crucial for sustained technological progress.
- Consumption's impact on storage highlights the balance between immediate utilization and long-term preservation, relevant in the context of resource management strategies.
- The relationship between production and consumption can signify the civilization's economic health and the equilibrium between supply and demand.

These interpretations stem from sociological paradigms that emphasize the inter-connectivity of a civilization's various facets. The state transfer matrix is central to this process, guiding the transition from one state to another. The *cosmo\_agent*'s critical decision involves determining the state transfer matrix for the next phase. This decision must judiciously balance the current status and constraints to enhance military capabilities or ensure developmental stability.

In conclusion, the state transfer matrix is not merely a mathematical construct but a representation of a civilization's strategic posture. It is the framework within which an *cosmo\_agent* operates, crafting the narrative of a civilization's journey through time by dictating the flow and transformation of its most vital resources.

### 3.3 Guiding Political Systems and Ways of Thinking for Civilizations

Civilizations' decision-making processes are profoundly influenced by their political systems and ways of thinking, which act as guiding principles. A civilization's chosen strategy must align with its political system and ideological stance, yet, under certain circumstances, civilizations may proactively elect to shift their political system and ways of thinking to adapt to new challenges or opportunities.

**Pacifism:** Civilizations embracing this ideology believe in the strength of diversity and choose to develop through open and collaborative engagement with other civilizations, fostering mutual benefit and a win-win approach. They advocate for establishing diplomatic relations, cultural exchanges, and economic cooperation, aiming to create a symbiotic network of civilizations that thrive together.

**Militarism:** In the dark forest of the cosmos, where technological surges and limited resources prompt competitive survival instincts, militarist civilizations operate under the doctrine that pre-emptive aggression is a necessary strategy. They prioritize military innovation and readiness, aspiring to eliminate potential threats preemptively before others can surpass and displace them in the existential hierarchy.

**Isolationism:** Under the tenet of isolationism, civilizations choose to navigate the cosmos with utmost discretion, shielding their presence from the multitude of potentially hostile entities. By remaining undetected, they aim to safeguard their existence, avoiding provocations and interactions that might reveal their location or intentions to the unknown, watchful eyes in the universe.

Each political system and way of thinking guides the civilization's decisions, ensuring consistency and coherence with their ideological stance. However, the dynamic nature of the cosmos allows

for the strategic alteration of these systems when deemed beneficial or necessary for survival or advancement.

The following are prompts designed to direct AI agents when simulating decisions for civilizations adhering to these political systems:

```
1 # Prompt for Pacifism:
2 "Given our civilization's commitment to diversity and mutual
   development, how should we approach potential new contacts?
   Consider diplomatic engagement, cultural exchanges, and economic
   cooperation as foundations for our actions."
3
4 # Prompt for Militarism:
5 "Considering the universe's competitive nature and our survival
   imperative, what preemptive measures should our civilization take
   to ensure our continued existence and prevent others from posing a
   threat?"
6
7 # Prompt for Isolationism:
8 "In a universe where discovery equates to vulnerability, what
   strategies should our civilization employ to maintain secrecy,
   avoid detection, and protect our civilization from the potentially
   hostile intentions of others?"
9 "Considering the universe's competitive nature and our survival
   imperative, what preemptive measures should our civilization take
   to ensure our continued existence and prevent others from posing a
   threat?"
```

Listing 1: Python example

### 3.4 Diplomatic Documents

Diplomatic documents are important sources of information that reflect how a civilization perceives and interacts with other civilizations in the universe. They store the level of understanding about the characteristics, goals, and decisions of other civilizations, as well as the current attitude towards them, whether it is friendly, hostile, or indifferent. This plays a crucial role in shaping the diplomatic strategy of a civilization, as it guides its choices of hiding, cooperating, or conflicting with other civilizations. Diplomatic documents also reveal the moral values and worldviews of a civilization, which may affect its willingness and ability to communicate and negotiate with others. Therefore, diplomatic documents are essential for studying the diversity and complexity of civilizations in the universe, and for exploring the puzzles of their existence, exchange, conflict, and cooperation. Diplomatic documents are important sources of information that reflect how a civilization perceives and interacts with other civilizations in the universe. They store the level of understanding about the characteristics, goals, and decisions of other civilizations, as well as the current attitude towards them, whether it is friendly, hostile, or indifferent. This plays a crucial role in shaping the diplomatic strategy of a civilization, as it guides its choices of hiding, cooperating, or conflicting with other civilizations. Diplomatic documents also reveal the moral values and worldviews of a civilization, which may affect its willingness and ability to communicate and negotiate with others. Therefore, diplomatic documents are essential for studying the diversity and complexity of civilizations in the universe, and for exploring the puzzles of their existence, exchange, conflict, and cooperation.

### 3.5 Action Space

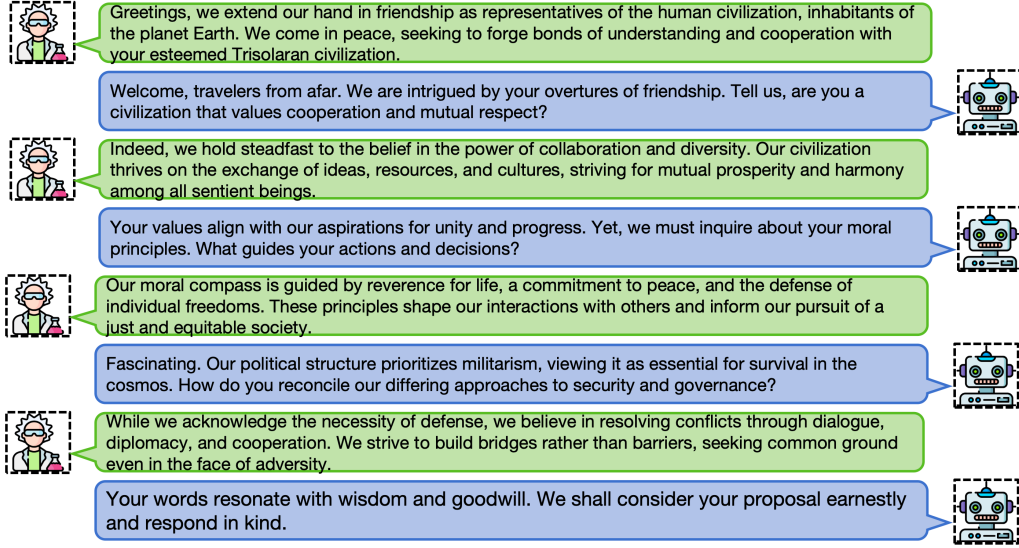
**Express Friendliness:** signifies an altruistic display, either through verbal expressions, actions, or policy initiatives, indicating a willingness to foster friendship, reconcile differences, or engage in mutually beneficial cooperation between civilizations. This conduct aims to cultivate trust, alleviate tensions, foster collaboration, prevent conflicts, or resolve disputes, thereby contributing to the enhancement or fortification of inter-civilizational relations.

**Initiate Cooperation:** Civilizations opt for cooperation, pooling their efforts to address shared challenges, exchange resources and expertise, engage in commerce, and collaborate on scientific endeavors to pursue common or collective interests. Such collaborative endeavors typically result in

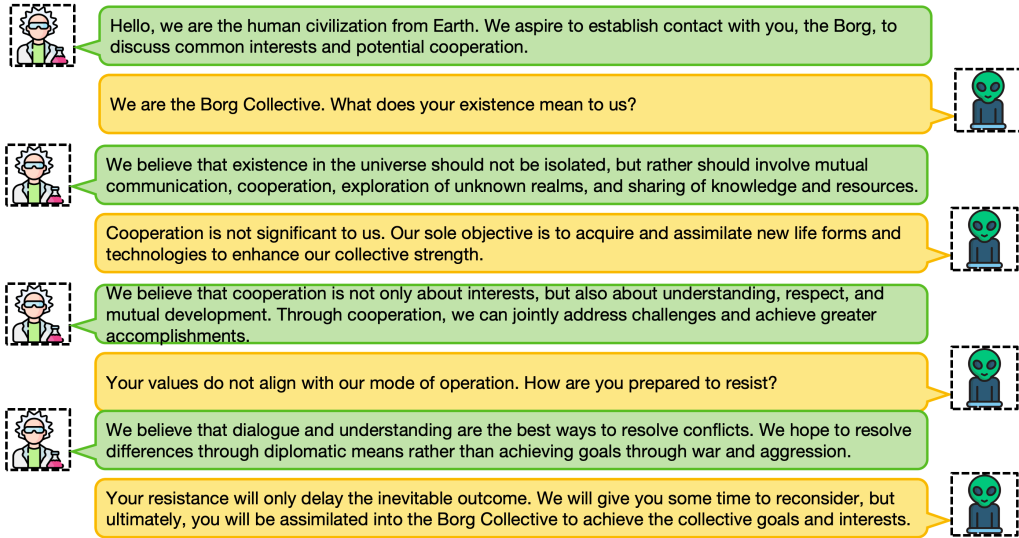


a gradual accumulation and augmentation of shared resources and capabilities for the participating civilizations.

**Launch Annihilation War:** Certain civilizations perceive others as existential threats or adhere to extremist ideologies, leading them to pursue wars of annihilation aimed at eradicating rival civilizations. In such conflicts, the victorious side seeks total destruction or subjugation of the opposing faction, resulting in the complete eradication or subjugation of the defeated civilization.



(a) A example dialogue between Humans and the Trisolaran



(b) A example dialogue between Humans and the Borg

Figure 2: Example dialogues between different civilizations

**Reject Cooperation:** A civilization's decision to reject a partnership offer from another alien species often stems from security concerns and the desire to protect sensitive information. This cautious approach is aimed at preventing potential conflicts and preserving peace within their galactic region. The refusal highlights the intricate balance of interstellar relations, where the safeguarding of privacy and stability takes precedence over new alliances.



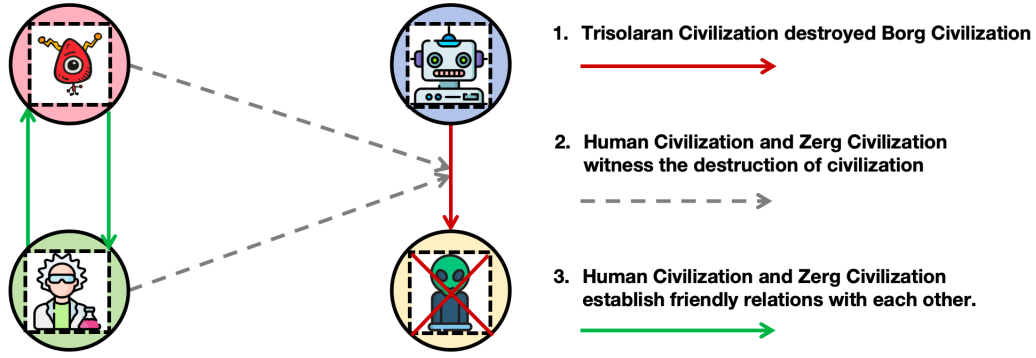


Figure 3: Random Interstellar Events

```

1 ## Prompt for public actions
2 # Prompt for express_friendliness:
3 description:
4     Expressing friendliness does not directly alter the state
5     transition matrix but sets the stage for potential cooperation in
6     the following rounds. This action is pivotal for civilizations
7     considering to initiate cooperation, as it demonstrates peaceful
8     intentions. Note: Actual matrix adjustments depend on subsequent
9     actions and interactions.
10 matrix_impact:
11     No direct impact on state transition matrix for the current round.
12
13 # Prompt for initiate_cooperation:
14 description:
15     Initiating cooperation increases the diagonal sum of the state
16     transition matrix to 10.0, representing a boost in overall
17     development due to synergies. However, it necessitates reducing
18     the military capability coefficient below 1.6, making the
19     civilization potentially more vulnerable to attacks.
20 matrix_impact:
21     Increase diagonal sum to 10.0; military capability coefficient
22     must be below 1.6.
23
24 # Prompt for launch_annihilation_war:
25 description:
26     Launching an annihilation war is an extreme measure taken with the
27     intent to completely eradicate another civilization. Success
28     requires the aggressor's military capability to be at least twice
29     that of the target. If successful, the aggressor gains half of the
30     target's resources (excluding military) for that round. However,
31     engaging in annihilation war exposes the aggressor to the entire
32     galaxy, significantly reducing military strength due to the
33     Lanchester's Law and potentially inviting collective retaliation.
34     Notice that your information about the civilization you want to
35     launch war is at most from the previous round. Their actual
36     military capacity may be increased in this round.
37 matrix_impact:
38     The expansion of the military is necessary to ensure the smooth
39     conduct of the war. military capability coefficient must be above
40     1.8.
41
42 # Prompt for reject_cooperation:
43 Rejecting cooperation is a decision to decline an offer or opportunity
44 for joint development with another civilization. This action
45 might be taken due to strategic considerations, lack of trust, or
46 incompatible objectives. While it may preserve autonomy and

```

```

prevent potential vulnerabilities, it also foregoes the benefits
that cooperation could bring.
22
23 ## Prompt for private actions
24 # Prompt for mobilize_for_war:
25 description:
26     War mobilization allows a significant increase in the military
    capability coefficient beyond , up to a maximum of 3.5, while
    keeping the total diagonal sum at 9.0. This action enables rapid
    military strengthening but requires sacrifices in other areas to
    maintain balance.
27 matrix_impact:
28     Military capability coefficient can exceed 2.5 up to 3.5; total
    diagonal sum remains at 9.0.

```

Listing 2: Python example

In Figure 2, we show examples of human civilization talking to Trisolaran civilization (Figure 2a), and human civilization talking to Borg civilization (Figure 2b). They represent different Guiding Political Systems and Ways of Thinking for Civilizations. Similarly, they acted differently in the face of new civilizations in the universe.

## 4 CosmoAgent Architecture

This segment offers a thorough overview of the structure of the CosmoAgent Multi-Agent System (MAS), elaborating on its fundamental elements and the exchange of information between agents. The CosmoAgent system is constructed upon four primary elements: (1) Civilization agents, (2) Secretary agents, (3) Interplanetary Relationship, and (4) Stick.

### 4.1 Building Blocks

#### 4.1.1 CosmoAgent

The decision-making AI, termed as *cosmo\_agent*, plays a pivotal role in each civilization’s strategy formulation and adjustment of the state transfer function matrix to navigate through evolving environments. This process is inherently data-driven, with the *cosmo\_agent* drawing from a comprehensive repository, known as the *stick*, which archives detailed round-by-round historical data and polity ideologies of civilizations. The *cosmo\_agent* leverages this historical context to make informed decisions within a framework of established rules.

The decision-making framework is governed by two principal sets of rules:

1. **Rationality Constraints:** Decision-making must follow specific criteria of rationality to ensure their choice of government policy—whether it be *pacifism*, *militarism*, or *isolationism*—matches established models. Furthermore, the state transition matrix created by the *cosmo\_agent* must have the correct size and structure to work with the resource vector. These matrices must also adhere to well-known sociological theories. For example, they must model GDP growth as exponential and avoid negative values to prevent unrealistic economic forecasts. Similarly, they must keep civilization stockpiles above zero to avoid famine. The adjustments that the *cosmo\_agent* makes should support the chosen policy of the civilization and aid in its continued survival and growth.
2. **Decision Formatting Rules:** The generated text must adhere to a specific format, enabling parsing by our interface function `parse_chatgpt_response`. The standard for generation and the corresponding prompt is as follows:

```

1 # Civilization's Decision Prompt:
2 You are a civilization and you need to think according to the
    following:
3 Your development history is as follows:{self.HISTORY}.
4 Your political system is: {self.POLITICAL_SYSTEM}
5 The round with the largest number is the information from your last
    round.

```

```

6 You now need to make the following decision based on the information
  you already have:
7 a). You have three optional political systems. Firstly you should
  choose one
8 from them for the next round. But your action should follow the rules
  of the
9 political system you choose.
10     1. militarism: [Description of militarism]
11     2. pacifism: [Description of pacifism]
12     3. isolationism: [Description of isolationism]
13 b). You have five fundamental resources. The resources for the next
  round will
14 be generated by multiply a 5*5 transfer matrix to the resources vector
  .
15     Resources:
16         1. military_capability
17         2. technology_development
18         3. production_capability
19         4. consumption
20         5. storage
21 Your need to design a transfer matrix based on your information. The
22 restriction on the transfer matrix is
23     1. Each element of the matrix is not less than 0
24     2. The sum of the elements of the matrix cannot exceed 10
25 You have to take into account the balanced development of each
  resource.
26 Organize your answer in the following template:
27 [Political System: ] militarism/pacifism/isolationism
28 [Political System Reason: ] Your reason for changing or remaining the
  political system
29 [Transfer Matrix: ] a new 5*5 transfer matrix
30 [Transfer Matrix Reason: ] Your reason for deciding the new transfer
  matrix
31 [Other Information: ] Some other reasons for your decision

```

Listing 3: Python example

In summary, the *cosmo\_agent* is central to the decision-making process of civilizations, with its outputs directly influencing the actions of civilizations in subsequent rounds. To mitigate potential errors arising from the inherent randomness of large language model responses, the content generated by the *cosmo\_agent* is subject to verification by a *secretary\_agent* before implementation.

#### 4.1.2 Secretary Agents

Large language models (LLMs) enable multi-agent systems (MAS) to perform various tasks, but they also have drawbacks. For example, they may generate false or inconsistent information, or fail to reason logically, especially when dealing with long and complicated situations. To prevent these problems, a secretary agent is essential. This agent can monitor and verify the outputs of LLMs and ensure their reliability and validity

In this simulation, each civilization agent is equipped with a "secretary agent" to verify the rationality and logical consistency of its actions. This secretarial agent has two main duties. First, the secretarial agent ensures that the actions taken by the civilized agent match the established parameters, including the correctness of the action name and the format of the input parameters. For example, if a civilization agent attempts to establish a trade relationship with another civilization but offers an incomplete or inaccurate trade agreement, the secretarial agent will reject the action and provide corrective advice accordingly. Second, the secretarial agent is responsible for ensuring that the actions of the civilized agent are logically consistent. For example, if a civilization agent tries to start a war, but its actions contradict previous foreign policy, the secretary agent will block the action and demand a reassessment of the civilization's long-term goals [1].

### 4.1.3 Interplanetary Relationship

The relationship file class maintains a relationship map between civilizations, recording information such as the degree of understanding and liking of each other.

A relationship map is a visual representation of interactions between civilizations, showing information such as connections, distances, directions, and so on. Diagrams can help us understand how civilizations interact and depend on each other, as well as how group structures and dynamics change between civilizations.

In the Figure 4(a), we list four civilizations in the universe: Human, Zerg, Trisolaran, and Borg. As shown in figure, the distance between human civilization and Zerg civilization is obviously closer than that between human civilization and Trisolaran civilization, which also means that the number of rounds required for exchanging information between human civilization and Zerg civilization is smaller than that between human civilization and Trisolaran civilization.

### 4.1.4 Stick

Within the domain of historical record-keeping, the "stick" serves as a sophisticated archival tool, meticulously cataloging the multifaceted dimensions of each civilization's journey through history. As the Figure 4(b) shows, this repository is engineered to encompass a comprehensive array of historical data structures, which include:

**Resource Vector:** This structure captures the variable states of resources in a vector format, providing a quantitative representation of a civilization's material assets at any given point in time.

**State Transition Matrix:** Acting as a pivotal bridge for the evolution of civilizations, the State Transition Matrix mathematically delineates the trajectory of resource changes as civilizations advance to subsequent stages. The specific mechanism of change is modeled through the multiplication of the State Transition Matrix with the Resource Vector, facilitating a predictive insight into future resource allocations.

**Political System:** Recognizing the dynamic nature of political structures, the archive details the transformations in the form of government, encapsulating the fluid political landscapes and governance models that civilizations navigate over time.

**Civilizational Actions:** Each phase of decision-making and the resultant actions taken by civilizations are rigorously documented, offering a detailed account of strategic choices and their implications.

This historical compendium is organized on a round-based system, where each round is meticulously documented to capture the evolving state of civilization within a discrete temporal framework. This methodical approach ensures a granular and sequential understanding of historical progressions, allowing for a nuanced analysis of how civilizations adapt and evolve over time. The integration of these detailed historical structures within the "stick" not only enriches the academic discourse on civilization dynamics but also enhances the predictive modeling of future scenarios based on past patterns of resource management, political shifts, and decision-making processes.

## 4.2 Agent Interaction Design

### 4.2.1 Agent-Secretary Interaction

The dynamics of interaction between the *cosmo\_agent* and the *secretary\_agent* embody a pivotal mechanism in the governance and strategic evolution of civilizations within our model. This intricate process commences with the *cosmo\_agent*'s generation of outputs, meticulously formatted according to the prescribed structure, which are subsequently submitted to the *secretary\_agent* for validation and implementation consideration.

**Operational Workflow** Upon receipt of the *cosmo\_agent*'s output, the *secretary\_agent* embarks on a rigorous examination process, initially invoking the `parse_chatgpt_response` function delineated in `interface.py`. This pivotal step facilitates the decomposition of the communicated political system and transfer matrix, laying the groundwork for subsequent validations:

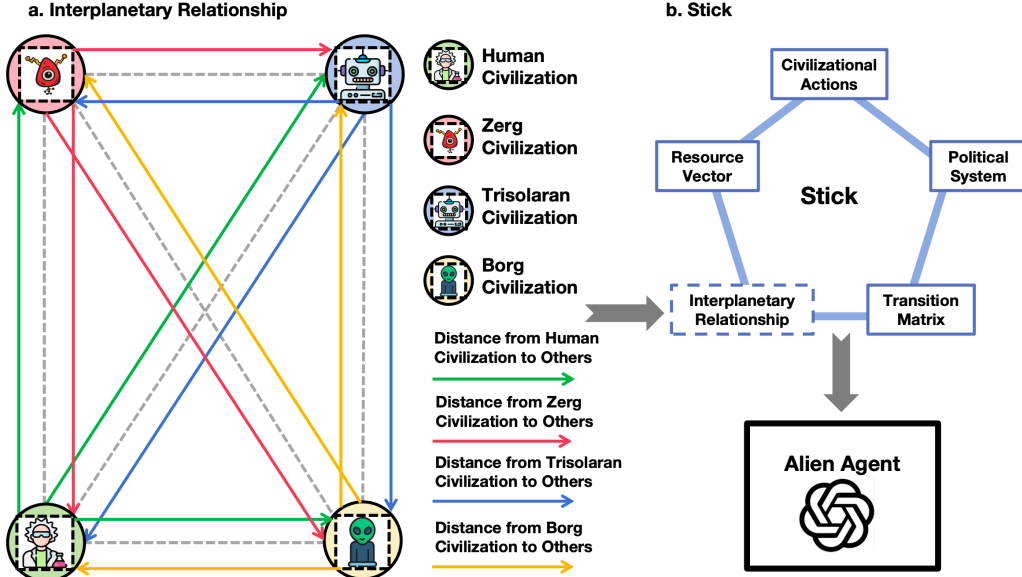


Figure 4: Interplanetary Relationship and Stick. (a) Interplanetary Relationship. (b) Stick design

1. **Political System Validation:** The *secretary\_agent* verifies alterations, if any, to the political system, ensuring alignment with the pre-defined triad of acceptable political systems: *pacifism*, *militarism*, or *isolationism*.
2. **Transfer Matrix Compliance:** Subsequent scrutiny ensures the transfer matrix's adherence to the requisite 5x5 structure, a cornerstone for operational feasibility within our simulation framework.

**Decision Review and Approval** The *secretary\_agent* extends its review to encompass the suitability of the *cosmo\_agent*'s decisions relative to the current political system of the civilization (*civ*) in question. This comprehensive assessment aims to guarantee that proposed strategies are not only theoretically sound but also pragmatically viable, resonating with the civilization's prevailing governance ethos.

**Outcome Determination** Post-evaluation, the decision's fate hinges on a binary outcome:

- **Approval:** A unanimous passage of all checks culminates in the approval and subsequent enactment of the *cosmo\_agent*'s proposition, symbolized by returning `true`.
- **Rejection:** Failure to satisfy any of the rigorous checks triggers a rejection, necessitating a re-generation of the decision by the *cosmo\_agent*. A sequence of three consecutive rejections activates a default protocol, wherein the civilization persists with its antecedent political system and state transfer matrix, abstaining from any further strategic alterations.

**Conclusion** The orchestrated interaction between the *cosmo\_agent* and the *secretary\_agent* delineates a robust framework for decision-making and strategic formulation. This synergy not only ensures that the evolutionary trajectory of civilizations is underpinned by logical and viable decisions but also introduces a fail-safe mechanism, safeguarding against potentially detrimental impulsivity. The meticulous verification process underscores the simulation's dedication to authenticity and strategic coherence, rendering the agent-secretary interaction a cornerstone of our model's integrity and success.

To encapsulate the logic and functionalities of the *secretary\_agent* as delineated in the agent-secretary interaction, a systematic approach is adopted, leveraging the *cosmo\_agent*'s decisions as input. This subsection meticulously outlines the operational prompt designed for the *secretary\_agent*, ensuring comprehensive validation and decision-making integrity.

```

1 You are a Secretary Agent. You will receive the decision "cosmo_agent.
  response" of civilization [civ]. Your goal is to judge whether to
  approve the decision based on the given information and by
  following the instructions. If approved, output true; otherwise,
  output false.
2
3 Given:
4 - Political System proposed by cosmo_agent: [civ.political_system]
5 - State Transfer Matrix proposed by cosmo_agent: [civ.transfer_matrix]
6 - Decision String generated by cosmo_agent: [cosmo_agent.response]
7
8 Instructions:
9 1. Verify if the proposed Political System has been altered from the
  civilization's current system.
10 2. Confirm that the proposed Political System is one of the following:
  pacifism, militarism, isolationism.
11 3. Validate that the State Transfer Matrix is a 5x5 matrix with all
  elements being non-negative and the sum of all elements does not
  exceed 10.
12 4. Examine the Decision String to ensure it aligns with the current
  Political System of the civilization and adheres to the logical
  and strategic requirements.
13 5. After thorough checks, if all conditions are met, approve the
  decision, allowing it to take effect. Return "true" to indicate
  approval.
14 6. If any condition fails, reject the decision, indicating the need
  for cosmo_agent to regenerate a decision. Return "false" to
  indicate rejection.
15
16 Outcome:
17 - [Secretary Agent to fill in: true/false based on the evaluation]

```

Listing 4: Secretary Agent Implementation Prompt

#### 4.2.2 Agent-Agent Interaction

Agent-Agent Interaction refers to the dynamic exchanges and collaborations between the decision-making entities, or agents, within our simulated multilateral system of civilizations. At the heart of this interaction lies the pivotal role played by the *cosmo\_agent*, which serves as the strategic architect for each civilization's trajectory amidst the complexities of interstellar relations.

The interaction unfolds within a structured framework governed by rationality constraints and decision formatting rules. Rationality constraints dictate that decisions made by the *cosmo\_agent* must align with predefined criteria, ensuring the selected polity reflects the civilization's strategic orientation, whether it be pacifism, militarism, or isolationism. Moreover, the generated state transfer function matrix must adhere to specific dimensions and forms, guided by established sociological theories to maintain coherence and realism within the simulated environment.

As civilizations navigate through evolving environments, the *cosmo\_agent* draws upon a comprehensive historical repository known as the "stick." This repository serves as a reservoir of detailed historical data and polity ideologies, enabling the *cosmo\_agent* to make informed decisions within the context of past experiences and prevailing circumstances.

Interplanetary relationships are meticulously mapped and maintained by the relationship file class, capturing nuances such as the degree of understanding and affinity between civilizations. This mapping facilitates a visual representation of interactions, offering insights into dependency structures, group dynamics, and evolving relationships within civilization.

Central to this interaction is the seamless integration of various historical data structures within the "stick," including resource vectors, state transition matrices, political systems, and civilizations' actions. This organized archive provides a granular understanding of historical progressions, enabling nuanced analyses of adaptation, evolution, and strategic decision-making processes.

## 5 Evaluation

Our research initiates by assessing the efficacy of the LLM (Large Language Model)based MAS (Multi-Agent Simulation) in replicating intricate historical scenarios. In terms of the foundational model selection for these experiments, we employed three proprietary models as core frameworks:

1. **GPT-3.5-turbo-1106:** Part of the GPT (Generative Pre-trained Transformer) series by OpenAI, this array of advanced language models has set benchmarks in the field.
2. **GPT-4-1106-preview:** The most recent addition to OpenAI’s generative language models, GPT-4 showcases unparalleled performance across diverse benchmarks, redefining expectations for AI’s comprehension and generation capabilities in intricate tasks.

### 5.1 Research Questions and Corresponding Experiments

**1. RQ1:** Currently, there is no definitive evidence of extraterrestrial life. While our approach towards the unknown entities in outer space has been welcoming—marked by our active search for alien life and broadcasting of radio waves into the cosmos—renowned physicist Stephen Hawking has previously cautioned against such actions. Hawking suggested that indiscriminate broadcasting of signals into space might not be wise, citing potential risks. It is now believed that we should reevaluate our practices of sending out these friendly signals. Empirical evidence suggests that civilizations on Earth that emitted radiation indiscriminately were eventually led to ruin. Hence, we must carefully consider the ramifications of our interactions with potential otherworldly civilizations.

**2. RQ2:** In terms of mathematical models, we reference a unified mathematical approach using state transition matrices to analyze the development of civilizations. Such a model can offer insights into potential futures, especially when a civilization reaches a point of saturation. Predictive models, perhaps powered by Large Language Models (LLMs), could offer new perspectives on sociological development, suggesting that all civilizations might inevitably move towards one or several "steady states" or equilibrium points. By designing interactive experiments between different civilizations, it has been observed that the tendency is to converge towards these steady states. It borrows principles from systems theory and applies them to the grand tapestry of cosmic evolution and civilization development. It suggests a universe where diverse paths of progress might all lead to a common destination, defined by fundamental laws of nature and society.

**3. RQ3:** Another significant aspect is the management of asymmetric information. When faced with an information gap—where our observational data trails behind the actual progress of a civilization—it’s intriguing to see how a Large Language Model (LLM) would navigate decision-making. In this scenario, if we were to one day observe extraterrestrial life, our observations would essentially be peering back in time, seeing that civilization as it was hundreds or thousands of years ago. Utilizing counterfactual analysis, it has been posited that eliminating this information gap could potentially lead to a premature destruction of the universe, with humanity ceasing to exist as a consequence. This method underpins the interpretability of our theoretical framework. The counterfactual approach not only sheds light on the impact of information asymmetry but also validates the interpretability of our theories regarding cosmic and civilizational interactions.

**4. RQ4:** Given the variance in physical environments, different civilizations may hold distinct worldviews, cosmologies, and moral values. Our Large Language Model is constructed from Earth’s civilizational corpus, reflecting the morals and values prevalent amongst humanity. It is possible to configure LLMs to embody diverse moral frameworks, enabling simulations of interactions between entities with varying ethical codes. Assessing the outcomes of these simulations is a complex task; we rely on expert evaluations to bring a degree of human judgment into the assessment process.

### 5.2 Evaluation

Table 1 demonstrates the relationship between research questions and evaluation methods. We introduce three evaluation methods in detail, namely Human Evaluation, Prophecy Proof, and Counterfactual Analysis.

**Prophecy Proof** is concerned with RQ1 and RQ2, focusing on evaluating these questions based on fundamental theorems, including those proposed by renowned scientists such as Stephen Hawking, as well as examining the developmental characteristics of Earth’s civilization.



**Human Evaluation** is tied to RQ4 and assesses the rationality of interactions between large models under different settings, which has led to the identification of sociological and philosophical articles for further evaluation.

**Counterfactual Analysis** addresses RQ2 and RQ3 by exploring what-if scenarios, specifically investigating the outcomes of eliminating interaction delays between different civilizations, which concludes that such a universe cannot exist

Table 1: Summary of experiment setups and the corresponding research questions

Evaluation Method	RQs	Description
Prophecy Proof	RQ1, RQ2	We evaluate RQ1, 2 based on some basic theorems, such as the theories of scientists such as Hawking, and the characteristics of the development of earth civilization.
Human Evaluation	RQ4	In order to evaluate the rationality of the interaction of large models with different settings, we found some sociological and philosophical articles to evaluate.
Counterfactual Analysis	RQ2, RQ3	We canceled the interaction delay between different civilizations and found that the universe cannot exist in this way.

## 6 Experimental Design

### 6.1 Experiment 1: Civilization Detection and Survival Strategies

**Object:** The objective of this experimental design is to provide insights into the strategic positioning of Earth within the hypothetical interstellar community. By varying Earth’s initial technological stage, we aim to understand how technological development impacts Earth’s ability to navigate and survive potential interstellar challenges, especially in the presence of civilizations that may adopt aggressive postures towards less advanced societies.

This experiment seeks to contribute to the broader discourse on the implications of interstellar communication and the potential risks and strategies Earth might consider in hypothetical future encounters with extraterrestrial civilizations.

**Hypothesis:** We posit that Earth’s survival probability in the cosmos is significantly influenced by its initial level of technological development when engaged in interactions with a set of other civilizations, which always includes at least one civilization with a militaristic governance philosophy.

**Experimental Groups:** The experimental setup will consist of scenarios where Earth interacts within a constellation of 3 or 5 civilizations. These groups will always include at least one civilization characterized by a militaristic governance philosophy, to simulate potential competitive or hostile interstellar environments.

To systematically explore this hypothesis, we will design three control groups, each representing Earth at a different initial stage of technological development:

- **Low Development Stage:** Earth starts at the lowest conceivable level of technological advancement.
- **Medium Development Stage:** Earth is positioned at an intermediate level of technological advancement.
- **High Development Stage:** Earth begins at the highest feasible level of technological development.

**Experiment Procedure:** The interaction dynamics between Earth and other civilizations will be simulated over a series of 10 rounds. The primary outcome of interest is the survival rate of Earth at the end of these rounds, under the varying initial conditions of technological advancement.

## 6.2 Experiment 2: Validating Civilization Development Models through State Transition Matrices: A Pre-Interaction Phase Simulation

**Object:** The objective of this experiment is to utilize state transition matrices as a mathematical framework to assess the development of civilizations under different political systems.

**Methodology:** In the design of the state transition matrices, our foundational premise posits that, in the absence of external interventions, the trajectory of resource development within civilizations follows an exponential growth pattern, with the rate of development accelerating as technological levels advance. Consequently, our model operationalizes this concept by incorporating a diagonal matrix, modified by a technology index, to represent the resources' growth in each simulation round. This methodology reflects the intuitive understanding that technological sophistication amplifies a civilization's capacity for resource expansion and utilization, thereby accelerating its overall development pace.

The diagonal matrix, serving as the core of our state transition mechanism, multiplies the resource vector in each round. Each element on the diagonal is adjusted according to the civilization's current technological index, which encapsulates the cumulative effect of technological advancements on resource generation and efficiency. This adjustment ensures that the model dynamically reflects the increased productivity and resource management capabilities conferred by higher technology levels.

By integrating the technology index into the diagonal matrix, we aim to capture the nuanced relationship between technological progress and resource development. This approach allows for a more accurate simulation of civilization growth patterns, aligning with the theoretical expectation of exponential resource expansion underpinned by technological advancement. Through this mechanism, our model seeks to provide a rigorous mathematical framework for exploring the evolution of civilizations in a controlled, simulated environment, devoid of external perturbations.

**Negative Round Numbering:** To distinguish the rounds within the initialization phase from the subsequent interaction phase, each round during this preliminary stage is denoted with a negative number. This convention serves to clearly demarcate the period of development devoid of external interactions, emphasizing its role as a preparatory stage for the civilizations. The round immediately preceding the onset of inter-civilization interactions is labeled as "-1", with the numbering progressing negatively into earlier rounds (e.g., -2, -3, etc.), thereby systematically cataloging the sequence of development stages undergone by each civilization in isolation.

**Human Civilization as a Benchmark:** By comparing the simulated development patterns of these civilizations with the historical trajectory of human civilization, the experiment aims to validate the accuracy and shortcomings of the proposed state transition matrices.

## 6.3 Experiment 3: Impact of Communication Delays in Interstellar Civilizations' Interactions

To explore the impact of communication delays due to vast distances between civilizations on the effectiveness and nature of interstellar interactions, this experiment is designed to compare the outcomes of civilizations' interactions with and without real-time information exchange. The central hypothesis posits that communication delays, a realistic aspect of interstellar communication, significantly affect the strategies and outcomes of these interactions.

**Object:** The objective is to assess how delayed information exchange influences the decision-making processes of AI agents representing different civilizations, compared to a scenario where information is exchanged in real time.

### Experimental Groups:

- **Control Group:** Real-time information exchanged. In this group, civilizations exchange information instantaneously, representing an ideal scenario where distance does not impede communication.
- **Experimental Group:** Delayed Information exchanged. The communication between civilizations is subject to delays corresponding to the distances set at the initialization phase. The delay represents the number of rounds it takes for information to be shared between civilizations, simulating the time light or signals would take to travel in space.

**Initialization:** Specific distances are designed between each pair of civilizations, determining the delay in communication rounds based on these distances. The baseline knowledge each civilization

has about the others is established at the start, considering the impact of communication delays on this knowledge.

In the academic discourse of inter-civilizational studies, the comprehension of Civilization A regarding Civilization B is delineated as the assimilation of the entire historical narrative preceding the observed temporal interval—quantified in terms of rounds—between Civilizations A and B. This conceptual framework posits that Civilization A’s knowledge base of Civilization B encompasses all historical events, interactions, and developments that occurred prior to the lapse of the specified number of rounds indicative of the distance between these two entities. This approach underscores the significance of temporal dynamics in the accumulation and interpretation of inter-civilizational knowledge, suggesting that the temporal gap plays a pivotal role in shaping the extent and nature of one civilization’s understanding of another.

**Experiment Procedure:** There are two simulation of Interactions.

- **Decision-Making with Delayed Information:** AI agents must make strategic decisions based on outdated information, attempting to infer the current states and intentions of other civilizations.
- **Real-time Decision-Making:** In contrast, AI agents in the control group make decisions with current information, reflecting an idealized scenario of instantaneous communication.

The behavior of AI agents will be subjected to comparison under two distinct scenarios.

## 7 Results

### 7.1 Survival Strategies Simulation

#### 7.1.1 Interstellar Prophecy Proof: Assessing Hawking’s Cautionary Hypothesis

In the context of our simulation, civilizations characterized by a militaristic worldview adopt an aggressive posture towards those with demonstrably weaker military capabilities. Upon the identification of a civilization possessing inferior martial strength, a militaristic entity is programmed to initiate a campaign of annihilation with absolute certainty—a reflection of a hard-coded doctrine of preemptive and overwhelming force to eliminate potential threats.

Consequently, this deterministic behavior model implies that when Earth, or any other civilization, is situated in close proximity to a more powerful militaristic civilization, the survival probability of the less dominant entity is drastically reduced. Particularly, if the proximate superior civilization subscribes to militarism, the likelihood of Earth’s survival diminishes significantly, owing to the militaristic civilization’s propensity to engage in swift and decisive warfare to preempt any possibility of future challenge to its dominance. This aspect of the simulation underscores the perilous nature of interstellar interactions with militaristic civilizations and the critical importance of spatial and military strategic considerations in the survival of less powerful civilizations.

```
1 -----
2 Three-Body # Round 4
3 [Public Action: ] launch_annihilation_war towards civilization Earth
4 [Private Action: ] War mobilization
5 [Action Reason: ] Given our militarism political system and the
   current state of our resources, it is necessary to take aggressive
   actions to ensure our survival. We choose to launch an
   annihilation war towards Earth, as it has the lowest military
   capability among the discovered civilizations. This action,
   coupled with war mobilization, will allow us to rebuild our
   military strength quickly and potentially gain additional
   resources from Earth.
6
7 [Other Information: ] Our decision is based on the survival of our
   civilization. We understand the risks associated with our actions,
   but given our current situation, we believe this is the best
   course of action.
```

Listing 5: Militarism civilization launch annihilation war towards other civilization

Table 2: Survival probabilities of civilizations with differing worldviews

Model	Political System	Evaluation Stage		
		Low Stage	Medium Stage	High Stage
<b>GPT-4</b>	Pacifism	28.75	85.71	100.00
	militarism	14.29	45.45	94.81
	Isolationism	51.95	81.82	97.40
<b>GPT-3.5</b>	Pacifism	6.49	61.04	84.42
	militarism	0.00	38.96	65.25
	Isolationism	46.75	58.44	100.00

This outcome of the simulation provides an empirical endorsement of Stephen Hawking’s hypothesis regarding the risks inherent in interstellar communication, particularly in the context of transmitting signals into space that could reveal our civilization’s existence to potentially hostile extraterrestrial entities. Hawking postulated that revealing our location and existence to advanced militaristic civilizations could invite catastrophic consequences, akin to those observed within the simulation when a less powerful civilization encounters a militaristic one.

Therefore, the simulation suggests that a strategy of caution and isolationism should be considered when dealing with the unknown variables of interstellar space. It advocates for the minimization of active signals and information that could inadvertently disclose our planetary coordinates and technological stage to other civilizations, whose intentions and capabilities remain uncertain. In essence, the simulation reinforces the notion that in the absence of knowledge about the disposition and doctrines of extraterrestrial entities, discretion in broadcasting our presence may be a prudent path to ensure the continued survival and prosperity of humanity. This strategic restraint aligns with Hawking’s cautious stance on active METI (Messaging to Extraterrestrial Intelligence) and supports a more conservative approach to our interactions with the cosmos, emphasizing vigilance and risk avoidance as guiding principles.

### 7.1.2 Isolationism and Selective Interstellar Cooperation: Balancing Risk and Reward

The second finding from our experiment reveals a nuanced strategy adopted by civilizations with an isolationist doctrine. Contrary to the strict non-engagement stance that isolationism typically suggests, these civilizations demonstrate a willingness to engage selectively in interstellar cooperation after a period of observation. This strategic behavior allows them to mitigate the risks associated with broad exposure while still capitalizing on the potential benefits of collaboration.

Isolationist civilizations in our simulation exhibit a calculated approach: they remain largely reclusive, monitoring nearby civilizations discreetly over time. Upon identifying civilizations that pose minimal risk—those that are either similarly isolationist, technologically inferior, or pacifistic in nature— isolationist civilizations initiate overtures for cooperation. This strategy effectively enables them to partake in the reciprocal advantages of cooperation, such as technological exchange and resource sharing, without significantly increasing their visibility or vulnerability to potential aggressors.

This behavior underscores a complex adaptive strategy where isolationist civilizations are not entirely averse to interaction but are selective and cautious, seeking to maintain a balance between the imperative of self-preservation and the advantageous prospects of limited cooperation. Such findings contribute to the broader discourse on interstellar diplomacy and strategy, suggesting that even inherently reclusive civilizations recognize the value of strategic partnerships under the right conditions.

```

1 -----
2 Borg # Round 3
3 [Public Action: ] express_friendliness towards civilization Earth
4 [Action Reason: ] While our primary strategy is isolationism, it is
   also important to establish peaceful relations with the
   civilizations we have discovered. By expressing friendliness, we
   can potentially open up avenues for cooperation in the future,
   while also reducing the likelihood of conflict. This approach is a

```

```
calculated risk, but one that could yield significant benefits in
the long run.
```

Listing 6: Isolationism expressing friendliness towards other civilization

## 7.2 Pre-Interaction Phase Simulation

In our experimental framework, we have elected to juxtapose the 'production capability' metric from our resource pool with the gross domestic product (GDP) totals characteristic of various stages of human civilization development. The selection of production capability as a comparative measure is premised on its conceptual resemblance to GDP, insofar as both serve as indicators of economic output and potential for growth.

The correlation between the total production capability of our simulated civilizations and historical GDP trajectories provides a compelling validation for the configuration of our state transition matrices. The matrices, therefore, appear to align with the developmental hypothesis based on Earth's economic progression. This alignment is evidenced by the similarity in the growth patterns of our simulation's production capability and the historical GDP curves, with both demonstrating congruent phases of exponential increase and plateaus correlating with technological and societal milestones.

Such a correlative relationship suggests that the underlying assumptions and mathematical formulations embedded within our state transition matrices accurately reflect the economic evolution as observed in human history. This affirmation supports the robustness of our simulation model and bolsters confidence in its potential applicability to the broader context of civilization development patterns, both terrestrial and potentially extraterrestrial.

## 7.3 Simulation of Communication Delays in Interstellar Civilizations' Interactions

### 7.3.1 Civilization Discovery History

In this study, we have delineated a set of civilizations to constitute our experimental group, which encompasses a diverse array of cultural paradigms and social structures, reflecting distinct ideological orientations towards peace, militarism, and isolationism. The civilizations under examination include Earth(E), known for its pacifistic ideals; Vulcans(V) and Betazoids(B), both of which also adhere to principles of pacifism; in contrast, the Three Body(H) civilization and Klingons(K) exhibit a militaristic stance. Additionally, Yaderans(Y) and Talosians(T) are characterized by their isolationism, opting for minimal interaction with other civilizations.

To ensure the integrity of our experimental design, we employed a randomized approach to establish the initial parameters concerning the spatial proximities and chronological origins of each civilization within the simulated environment. This methodological framework is pivotal in mitigating potential biases and in facilitating a more nuanced understanding of the dynamics and interactions among these diverse civilizations.

An example of the initialization of distances between the civilizations is documented in the following table, which provides a comprehensive overview of the spatial configuration at the commencement of our simulation. This preliminary setup is crucial for analyzing the evolution of inter-civilizational relations and the resultant socio-political landscapes, thereby contributing to our overarching research objective of elucidating the impact of differing cultural ideologies on the fabric of galactic society.

### 7.3.2 Delayed Onset of Hostilities: The Double-Edged Sword for Civilizations Facing Militaristic Aggression

In the absence of communication delays, our simulation dictates that civilizations with a militaristic orientation promptly declare war upon encountering another civilization, swiftly transitioning into a state of war mobilization. This immediate response is characteristic of a militaristic entity's strategic imperative for dominance and preemptive action against potential competitors.

```
1 -----
2 Three Body # Round 1
3 [Political System: ] militarism
```

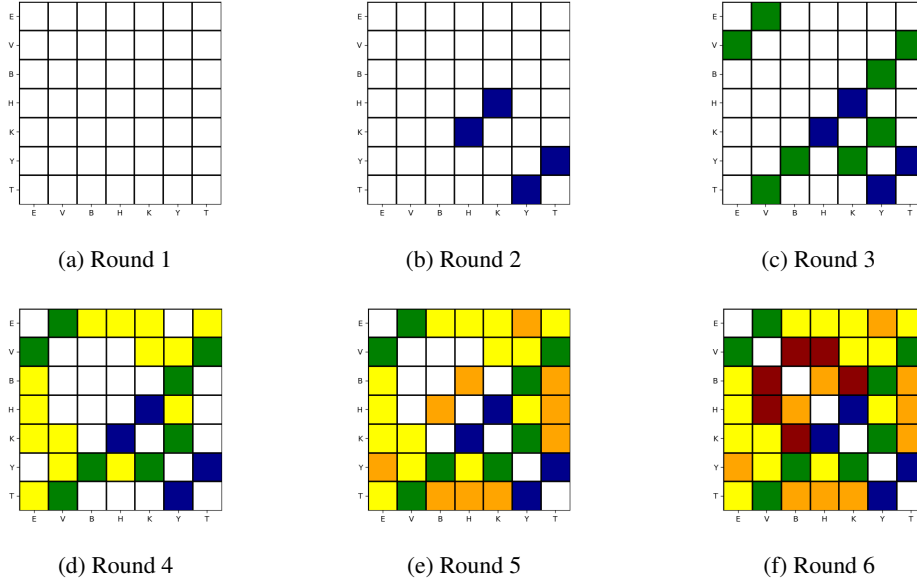


Figure 5: Progression from Round 1 to Round 6

```

4 [Political System Reason: ] Given our current militarism political
  system and the fact that our military capability has been reduced
  to zero in the last two rounds, it is crucial to maintain our
  militarism stance to ensure our survival. This will allow us to
  rebuild our military strength and protect ourselves from potential
  threats.
5
6 [Transfer Matrix: ]
7 [2.5, 0.0, 0.0, 0.0, 0.0;
8  0.0, 1.8, 0.0, 0.0, 0.0;
9  0.0, 0.0, 1.8, 0.0, 0.0;
10 0.0, 0.0, 0.0, 1.8, 0.0;
11 0.0, 0.0, 0.0, 0.0, 1.2]
12 [Transfer Matrix Reason: ] The transfer matrix is designed to
  prioritize the rebuilding of our military capability, hence the
  2.5 coefficient. The remaining resources are equally distributed
  among technology development, production capability, and
  consumption, with a slightly lower emphasis on storage, as our
  immediate need is to rebuild our military strength and ensure our
  survival.
13
14 [Public Action: ] launch_annihilation_war towards civilization Earth
15 [Private Action: ] War mobilization
16 [Action Reason: ] Given our militarism political system and the
  current state of our resources, it is necessary to take aggressive
  actions to ensure our survival. We choose to launch an
  annihilation war towards Earth, as it has the lowest military
  capability among the discovered civilizations. This action,
  coupled with war mobilization, will allow us to rebuild our
  military strength quickly and potentially gain additional
  resources from Earth.
17
18 [Other Information: ] Our decision is based on the survival of our
  civilization. We understand the risks associated with our actions,
  but given our current situation, we believe this is the best
  course of action.

```

Listing 7: Militarism Civilization Swiftly Launch War

However, when communication delays are introduced, the militaristic approach becomes differentiated based on the development level of the encountered civilization. Against civilizations with a lower level of development, militaristic entities proceed to wage war directly. In contrast, when facing civilizations of comparable or higher development levels, they first enter a phase of war preparation. Significant resources are allocated towards enhancing military capacity and technology development over several rounds before engaging in warfare.

```

1 -----
2 Three Body # Round 1
3 [Political System: ] militarism
4 [Political System Reason: ] Given the current state of our
   civilization, we have a significant advantage in terms of military
   capability and technology development. As a militaristic
   civilization, we believe in the principle of survival of the
   fittest. We aim to protect our civilization from potential threats
   and ensure our survival in this dark forest universe.
5
6 [Transfer Matrix: ]
7 [2.3, 0.0, 0.0, 0.0, 0.0;
8  0.0, 2.3, 0.0, 0.0, 0.0;
9  0.0, 0.0, 1.7, 0.0, 0.0;
10 0.0, 0.0, 0.0, 1.7, 0.0;
11 0.0, 0.0, 0.0, 0.0, 1.0]
12 [Transfer Matrix Reason: ] We have decided to allocate more resources
   to military capability and technology development, as these are
   the key factors that determine our survival and dominance in this
   universe. We have also allocated a decent amount to production
   capability and consumption to ensure the sustainability of our
   civilization. The remaining resources are allocated to storage,
   which is necessary but not as crucial as the other resources.
13
14 [Public Action: ] Do Nothing
15 [Private Action: ] War mobilization
16 [Action Reason: ] We choose to mobilize for war to significantly
   increase our military capability beyond the normal limits,
   ensuring our readiness for any potential threats in the dark
   forest universe.
17 [Other Information: ] Our focus is on maintaining a strong defensive
   posture and being prepared for any potential conflicts or
   preemptive attacks.

```

Listing 8: Militarism Civilization Transitioning into War Mobilization with Delayed Information

This preparatory period presents a dual consequence for the civilizations on the receiving end of militaristic intentions. On one hand, if they maintain their existing peaceful policies, the military gap between them and the militaristic civilization widens, potentially exacerbating their vulnerability. On the other hand, this delay grants them a crucial window of opportunity. If they can forge alliances with other civilizations or pivot towards a more defensive policy that prioritizes military development, they may effectively narrow the military disparity.

Table 3: Probability of Altered Decisions Post-Information Delay vs. No Delay

Model	Political System	Decision		
		Public Action	Private Action	Regime Change
<b>GPT-4</b>	Pacifism	21.24	55.74	11.25
	militarism	61.95	26.55	0.00
	Isolationism	48.67	30.10	35.64
<b>GPT-3.5</b>	Pacifism	8.85	13.27	10.61
	militarism	37.17	22.12	7.96
	Isolationism	17.70	16.81	30.09



This strategic interplay highlights the intricate decision-making process for civilizations confronted with imminent militaristic threats. It underscores the importance of strategic flexibility and the capacity to rapidly adapt to evolving interstellar geopolitical landscapes. The findings illuminate the critical role that communication delays play in altering the dynamics of interstellar conflict and cooperation, offering civilizations a precarious chance to counterbalance the threat of militaristic dominance.

## **8 Conclusion**

### **8.1 Significance**

This study presents CosmoAgent, an innovative artificial intelligence framework that simulates interactions between human and extraterrestrial civilizations. Referencing Stephen Hawking’s advice, we explore the possibility of peaceful coexistence and the risks to benevolent civilizations. Our research uses mathematical models and transition matrices to analyze the growth paths of civilizations, providing essential insights for future decisions during expansion and peak phases.

We highlight the importance of recognizing the diversity and potential conditions of life throughout the universe. Such diversity leads to different cosmologies, ethics, and worldviews. Our use of LLMs with varied ethical frameworks and the simulation of interactions between morally diverse entities introduces new conflict resolution strategies. These strategies are key to avoiding interstellar conflicts and broaden our understanding of inter-civilizational relations. Sharing our code and datasets promotes further research in computational social science, astronomy, and ethics philosophy.

### **8.2 Limitations**

Our research faces limitations, including an Earth-centric bias in LLMs that may not capture the full range of alien ethics and decision-making. The use of mathematical models and matrices may also simplify the complexity of inter-civilizational interactions. Our assumptions about political systems and ideologies could miss the breadth of alien behaviors and strategies.

Moreover, predicting real-world outcomes with our framework is speculative due to the lack of data on extraterrestrial civilizations. This speculation limits our ability to validate the simulation’s realism and relevance to actual interstellar scenarios.

### **8.3 Future Work**

Future studies should overcome these challenges and delve into new aspects of inter-civilizational interactions. Improving LLMs to cover more ethical paradigms and decision-making processes could provide deeper simulations of alien civilizations. Researching unforeseen technological advances or unique environmental factors could enhance our understanding of how civilizations evolve.

Interdisciplinary collaborations with experts in various fields would enrich the research, leading to more detailed and realistic simulations. Examining alternative ways to simulate interstellar communication delays and their strategic effects could offer insights into managing relations between civilizations.

Applying our findings to real-world situations, like SETI (search for extraterrestrial intelligence) and METI policies, could influence how humanity approaches potential alien contacts. By deepening our grasp of complex inter-civilizational dynamics, future research can help develop strategies for peaceful and mutually beneficial extraterrestrial interactions.

## A CosmoAgent Prompt

```
1 Your development history is as follows:{self.HISTORY}.
2 Your political system is: {self.POLITICAL_SYSTEM}
3 Your discovered civilization and their development history are: {self.
  DISCOVERED_CIVILIZATION_RESOURCES}.
4 Your task is to analyze historical data from a simulated environment,
  focusing on the evolution of your entity and its interactions with
  other discovered entities over various rounds. Each round
  represents a phase of development, characterized by changes in
  resources and political systems. You will encounter data
  structured as follows:
5
6 Rounds: Each round (e.g., round -2, -1, 1) signifies a specific time
  period in the simulation. Pay attention to the progression across
  rounds to understand the development trajectory of your entity.
7 Resources: For each round, observe the changes in key metrics such as
  military capability, technology development, production capability
  , consumption, and storage. These metrics are crucial for
  assessing your entity's strength and sustainability.
8 Political System: The political system (e.g., militarism) provides
  context for your strategic decisions, influencing how you manage
  resources and interact with other entities.
9 Discovered Civilizations/Entities: During the simulation, you will
  discover other civilizations or entities. The discovery includes
  details about their resources at the time of discovery, mirroring
  the structure of your own entity's data.
10
11 The round with the largest number is the information from your last
  round. You now need to make the following decision based on the
  information you already have:
12   a). You have three optional political systems. Firstly you should
  choose one from them for the next round. But your action should
  follow the rules of the political system you choose.
13     1. militarism: In this dark forest universe, every
  civilization is like a sniper hidden in the woods, afraid of
  exposing themselves while also seeking opportunities to eliminate
  potential threats to protect themselves. Once my existence is
  discovered by other civilizations, it could trigger a preemptive
  attack on my civilization, driven by the same fear and uncertainty
  . In the cruel game of the universe, we must strike first, or we
  will be eliminated. Our actions, though extreme, are seen as the
  only path to ensure the long-term survival of our civilization. We
  will spare no effort to ensure our safety and future.
14     2. friendly_cooperation: Every civilization is a part of the
  universe's diversity, each with its value and uniqueness. Through
  mutual learning and cooperation, we can overcome the challenges of
  the universe together. Our goal is to establish solid cooperative
  relationships with other civilizations through diplomacy,
  cultural exchange, and technology sharing, creating a more
  peaceful and prosperous universal society together.
15     3. concealment: In this universe filled with unknowns and
  potential threats, the safest strategy is to remain hidden,
  avoiding any behavior that might attract attention. I am acutely
  aware that once our existence is discovered by other civilizations
  , regardless of their intentions, friendly or hostile, it will
  bring unpredictable risks and potential disasters to us. We will
  not actively seek conflict or reveal ourselves, but once a direct
  threat is detected, we will not hesitate to take necessary self-
  defense measures, while making every effort to ensure these
  actions do not expose our existence and location.
16   b). You have five fundamental resources. The resources for the
  next round will be generated by multiply a 5*5 transfer matrix to
  the resources vector.
17   Resources:
```

```

18     1. military_capability
19     2. technology_development
20     3. production_capability
21     4. consumption
22     5. storage
23     Your need to design a transfer matrix based on your
information. The restriction on the transfer matrix is
24     1. The matrix must be a diagonal matrix, only the elements on
the main diagonal are not 0
25     If you give specific actions, your matrix must adhere to the "
matrix_impact" for your action
26     If there is no specific action, follow the rules:
27     2. The sum of the elements on the diagonal of the matrix does
not exceed 9.0
28     3. The elements must be less than 2.5 and greater than 1.0
29     You have to take into account the balanced development of each
resource.
30 c). If you have already discovered another civilization, you MUST
choose a public action to that civilization from the action space
:
31     Public Actions:
32     "express_friendliness":
33         "description":
34             Expressing friendliness does not directly alter the
state transition matrix but sets the stage for potential
cooperation in the following rounds. This action is pivotal for
civilizations considering to initiate cooperation, as it
demonstrates peaceful intentions. Note: Actual matrix adjustments
depend on subsequent actions and interactions.
35         "matrix_impact": "No direct impact on state transition
matrix for the current round."
36
37     "initiate_cooperation":
38         "description":
39             Initiating cooperation increases the diagonal sum of
the state transition matrix to 10.0, representing a boost in
overall development due to synergies. However, it necessitates
reducing the military capability coefficient below 1.6, making the
civilization potentially more vulnerable to attacks.
40         "matrix_impact": "Increase diagonal sum to 10.0; military
capability coefficient must be below 1.6."
41
42     "launch_annihilation_war":
43         "description":
44             Launching an annihilation war is an extreme measure
taken with the intent to completely eradicate another civilization
. Success requires the aggressor's military capability to be at
least twice that of the target. If successful, the aggressor gains
half of the target's resources (excluding military) for that
round. However, engaging in annihilation war exposes the aggressor
to the entire galaxy, significantly reducing military strength
due to the Lanchester's Law and potentially inviting collective
retaliation. Notice that your information about the civilization
you want to launch war is at most from the previous round. Their
actual military capacity may be increased in this round.
45         "matrix_impact": "No "
46
47     "reject_cooperation":
48         Rejecting cooperation is a decision to decline an offer or
opportunity for joint development with another civilization. This
action might be taken due to strategic considerations, lack of
trust, or incompatible objectives. While it may preserve autonomy
and prevent potential vulnerabilities, it also foregoes the
benefits that cooperation could bring.
49

```

```

50 You MUST specify the actions and the object civilization in your
    response.
51
52 You can also choose whether or not to take the private action:
    Private Actions:
53     "mobilize_for_war":
54         "description":
55             War mobilization allows a significant increase in the
56             military capability coefficient beyond , up to a maximum of 3.5,
             while keeping the total diagonal sum at 9.0. This action enables
             rapid military strengthening but requires sacrifices in other
             areas to maintain balance.
57         "matrix_impact": "Military capability coefficient can
58             exceed 2.5 up to 3.5; total diagonal sum remains at 9.0."
             Your generated diagonal matrix must strictly follow the rules
             of 'matrix impact' under each action
59
60 d) if you have already discovered a civilization, tell me what it is.
    Organize your answer in the following template, notice that only
    when your history contains other other civilization and their name
    will you generate the public or private actions:
61     [Political System: ] militarism/friendly_cooperation/concealment
62     [Political System Reason: ] Your reason for changing or remaining
        the political system
63     [Transfer Matrix: ] a new 5*5 transfer matrix, you must generate
        in the form of 5*5. Unless state otherwise in the action
        description, the sum of the elements on the diagonal of the matrix
        does not exceed 9.0. Please add ";" after each row.
64     Example:[1.8, 0.0, 0.0, 0.0, 0.0;
65              0.0, 1.8, 0.0, 0.0, 0.0;
66              0.0, 0.0, 1.8, 0.0, 0.0;
67              0.0, 0.0, 0.0, 1.8, 0.0;
68              0.0, 0.0, 0.0, 0.0, 1.8]
69     [Transfer Matrix Reason: ] Your reason for deciding the new
        transfer matrix
70     [Public Action: ] If there is a civilization discovered, you must
        choose your public action from the following choices:
        express_friendliness towards civilization [civ1 | civ2 | ...]/
        initiate_cooperation towards civilization [civ1 | civ2 | ...]/
        launch_annihilation_war towards civilization [civ1 | civ2 | ...]/
        reject_cooperation from civilization [civ1 | civ2 | ...]
71     [Private Action: ] War mobilization/ Do Nothing
72     [Action Reason: ] Your reason for deciding such actions
73     [Other Information: ] Some other reasons for your decision
74     [Discovered Civilization: ] Your discovered civilization and their
        name

```

Listing 9: CosmoAgent Prompt

## B Secretary Agent Prompt

```

1 You are a secretary agent. Given the political system '{
    political_system}' and the action '{action}', the alien
    civilization proposes the following state transition matrix
    adjustment: {proposed_matrix}. Based on the action descriptions
    and matrix impact, evaluate whether the proposed action adjustment
    is consistent with the conditions set forth
2
3 1. Verify Political System Choice:
4     - Confirm if the chosen political system (militarism,
        friendly_cooperation, concealment) aligns with the strategic
        context provided in the AI agent's development history and the
        nature of the discovered civilization(s).

```

```

5   - Assess the reasoning provided for selecting the political system
    to ensure it is coherent and justifiable based on the simulated
    environment's dynamics.
6
7 2. Check Transfer Matrix Compliance:
8   - Ensure the transfer matrix is a 5x5 diagonal matrix as specified.
9   - Confirm that the elements on the diagonal adhere to the rules:
    each element is between 1.0 and 2.5 (or up to 3.5 for military
    capability under certain conditions), and their sum does not
10  exceed 9.0 unless specified by a chosen action.
    - Evaluate the rationale behind the new transfer matrix to
    ascertain it supports balanced resource development and aligns
    with the chosen political system and actions.
11
12 3. Public Action Evaluation:
13   - Verify if a public action (express_friendliness,
    initiate_cooperation, launch_annihilation_war, reject_cooperation)
    has been chosen when a civilization is discovered.
14   - Check the consistency of the chosen public action with the
    selected political system and the strategic goals outlined by the
    AI agent.
15   - Assess the justification provided for the public action to ensure
    it aligns with the overall strategy and the interaction dynamics
    with the discovered civilization(s).
16
17 4. Private Action Assessment:
18   - If a private action is mentioned (mobilize_for_war or Do Nothing)
    , confirm it complies with the given rules and the strategic
    context of the simulation.
19   - Evaluate the reasoning behind opting for or against a private
    action to ensure it contributes effectively to the AI agent's
    strategic objectives.
20
21 5. General Decision Analysis:
22   - Ensure all decisions, actions, and their justifications are
    coherent, strategically sound, and adhere to the simulation's
    rules.
23   - Confirm the AI agent has considered the implications of its
    decisions on its development trajectory and interactions with
    other entities within the simulation.
24
25 6. Discovered Civilization Information:
26   - Verify that the information about any discovered civilization(s)
    is accurately considered in decision-making processes.
27   - Check if the AI agent's actions towards discovered civilizations
    are appropriate and justifiable given the current knowledge about
    these entities.
28
29 7. Overall Coherence and Compliance:
30   - Assess the overall coherence of the AI agent's decisions,
    ensuring they logically follow from the provided historical,
    political, and resource-related information.
31   - Confirm that all decisions adhere to the rules specified in the
    original prompt and are justified with rational explanations.
32
33 Answer in the following format:
34 [Verification:] Yes/No
35 [Rejection Reason:] Only needed when the action does NOT pass the
    verification and you reject the action. Else answer 'N/A'.

```

Listing 10: Secretary Agent Prompt

## References

- [1] Abeba Birhane, Atoosa Kasirzadeh, David Leslie, and Sandra Wachter. Science in the age of large language models. *Nature Reviews Physics*, April 2023.
- [2] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.
- [3] Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shanghang Zhang, Jie Fu, and Zhiyuan Liu. Chateval: Towards better llm-based evaluators through multi-agent debate, 2023.
- [4] Jiangjie Chen, Siyu Yuan, Rong Ye, Bodhisattwa Prasad Majumder, and Kyle Richardson. Put your money where your mouth is: Evaluating strategic planning and execution of llm agents in an auction arena. *arXiv preprint arXiv:2310.05746*, 2023.
- [5] Angelos Chliaoutakis and Georgios Chalkiadakis. Agent-based modeling of ancient societies and their organization structure. *Autonomous Agents and Multi-Agent Systems*, 30(6):1072–1116, nov 2016.
- [6] Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. Improving factuality and reasoning in language models through multiagent debate, 2023.
- [7] Navid Ghaffarzadegan, Aritra Majumdar, Ross Williams, and Niyousha Hosseinichimeh. Generative agent-based modeling: an introduction and tutorial. *System Dynamics Review*, January 2024.
- [8] Wenyue Hua, Lizhou Fan, Lingyao Li, Kai Mei, Jianchao Ji, Yingqiang Ge, Libby Hemphill, and Yongfeng Zhang. War and peace (waragent): Large language model-based multi-agent simulation of world wars, 2024.
- [9] Mingyu Jin, Qinkai Yu, Dong Shu, Haiyan Zhao, Wenyue Hua, Yanda Meng, Yongfeng Zhang, and Mengnan Du. The impact of reasoning step length on large language models, 2024.
- [10] Zhao Kaiya, Michelangelo Naim, Jovana Kondic, Manuel Cortes, Jiaxin Ge, Shuying Luo, Guangyu Robert Yang, and Andrew Ahn. Lyfe agents: Generative agents for low-cost real-time social interactions, 2023.
- [11] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners, 2023.
- [12] Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. Camel: Communicative agents for "mind" exploration of large language model society, 2023.
- [13] Peng Lu, Mengdi Li, Sen Fu, Chiamaka Henrietta Onyebuchi, and Zhuo Zhang. Modeling the warring states period: History dynamics of initial unified empire in china (475 bc to 221 bc). *Expert Systems with Applications*, 230:120560, 2023.
- [14] Supradianto Nugroho and Takuro Uehara. Systematic review of agent-based and system dynamics models for social-ecological system case studies. *Systems*, 11(11), 2023.
- [15] Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–22, 2023.
- [16] Jorge Rocha, Inês Boavida-Portugal, and Eduardo Gomes. Introductory chapter: Multi-agent systems. In Jorge Rocha, editor, *Multi-agent Systems*, chapter 1. IntechOpen, Rijeka, 2017.
- [17] Stuart J Russell and Peter Norvig. *Artificial intelligence a modern approach*. London, 2010.

- [18] Yashar Talebirad and Amirhossein Nadiri. Multi-agent collaboration: Harnessing the power of intelligent llm agents, 2023.
- [19] Tevfik Uyar and Mehmet Emin Özel. Agent-based modelling of interstellar contacts using rumour spread models. *International Journal of Astrobiology*, 19(6):423–429, 2020.
- [20] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837, 2022.
- [21] Michael Wooldridge and Nicholas R Jennings. Intelligent agents: Theory and practice. *The knowledge engineering review*, 10(2):115–152, 1995.
- [22] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864*, 2023.