

Article

Not peer-reviewed version

---

# Generalized Agent Theory from First Principles

---

[Liu Feng](#)\*, [Lv Benfu](#), [Liu Ying](#)

Posted Date: 4 June 2025

doi: 10.20944/preprints202506.0055.v1

Keywords: Generalized Agent Theory; Standard Agent Model; First Principles; Observer Agent; Capability Space Evolution



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Generalized Agent Theory from First Principles

Feng Liu <sup>1,3,\*</sup>, BenFu Lv <sup>1,2</sup> and Ying Liu <sup>2</sup>

<sup>1</sup> The Institute of Digital Economy, China Institute for Innovation and Development Strategy, 100091 Beijing, China

<sup>2</sup> School of Economics and Management, University of Chinese Academy of Sciences, 100190 Beijing, China

<sup>3</sup> Institute for Digital Brain Research, CST, 100089 Beijing, China; liufeng@idbr-cst.org

**Abstract:** To address the fragmentation in the definition of Agent and the profound challenges concerning the nature of intelligence, consciousness, and the observer-based unification of physics, this paper establishes Generalized Agent Theory from first principles. The cornerstone of this theory is Standard Agent Model, which conceptualizes Agent as a system possessing five indispensable functional modules: information input, output, dynamic storage, creation, and control, corresponding to a five-dimensional capability vector space. Based on Standard Agent Model, we systematically classify Agent, including Absolute Zero Agent (Alpha Agent), Omniscient and Omnipotent Agent (Omega Agent), and Finite Agent. and further detail 243 subtypes, while also exploring multi-agent relationships from the dimensions of perception, communication, and interaction. Subsequently, we introduce the “Polar Intelligent Field Model,” characterized by an Alpha-decay field and an Omega-enhancement field, to describe its evolutionary dynamics within capability space. The framework also proposes “Wisdom” (A metric for an agent's aggregate intelligence) as a key intrinsic property of Agent, central to its interaction with these fields. Generalized Agent Theory offers a new interpretation of intelligence and consciousness: intelligence is defined as the overall efficacy of Agent in achieving its evolutionary goals, while consciousness is identified as the control function and its operational process. This provides a principled criterion for assessing the state of consciousness in current artificial intelligence systems, such as large language models. Crucially, Generalized Agent Theory reconceptualizes the physical observer as a specialized form of Agent and posits that the Universe itself is a dynamically evolving Finite Agent. From this perspective, fundamental physical concepts such as subjectivity/objectivity, uncertainty, the nature of spacetime, and entropy are reinterpreted as relative categories contingent upon the capabilities of the observer-agent. The theory further posits that the differences among classical mechanics, relativity, and quantum mechanics stem from the disparate capability configurations of their respective implicit observers, thereby offering a novel path toward their unification. This research aims to provide a unified foundational framework to advance research in artificial intelligence, physics, and related disciplines.

**Keywords:** Generalized Agent Theory; Standard Agent Model; First Principles; Observer Agent; Capability Space Evolution

## 1. Introduction

The 21st century presents science with two fundamental challenges: the unification of general relativity and quantum mechanics in physics, and the exploration of the nature of intelligence and consciousness in artificial intelligence. These two ostensibly disparate fields possess profound intrinsic connections. Within the theoretical frameworks of physics, the observer occupies a pivotal position [1]—the measurement problem in quantum mechanics and the choice of reference frames in general relativity both underscore the observer's central role. Meanwhile, in intelligence science, Agent has become a focal Pole of research, with intelligence and consciousness understood as

intrinsic attributes of Agent [2]. In recent years, interdisciplinary research has indicated that the "observer" in physics can be regarded as a specific instance of Agent [3]. This realization suggests that constructing an Agent model based on first principles can not only advance intelligence science but may also offer new perspectives for unifying fundamental theories in physics. However, despite significant progress in Agent research, a unified theoretical framework for Agent rooted in first principles is still lacking [4]. Existing definitions of Agent predominantly remain at the level of functional descriptions, exhibiting fragmentation and lacking systematicity and internal consistency.

To address this challenge, this paper, adhering to first principles, aims to construct a unified theoretical model with universal explanatory power—Generalized Agent Theory (GAT)—by starting from the most fundamental constituents and operational mechanisms of Agent. Central to this theory is the proposal of Standard Agent Model (SAM), which strives for 'maximal functional optimality and universal explanatory power' under the premise of 'structural simplicity'. Standard Agent Model posits that any Agent is composed of five irreducible fundamental functional modules: Information Input (In), Information Output (Out), Dynamic Storage (DS), Information Creation (Cr), and Control (Con). Inspired by classical computational models such as the von Neumann architecture, this model particularly emphasizes Information Creation as a key capability distinguishing Agent from traditional computational systems, and more precisely integrates the 'computation' function into the unified concept of Dynamic Storage. The magnitudes of these five fundamental functions of Agent constitute its five-dimensional capability vector space.

Based on Standard Agent Model and its defined five-dimensional capability vector, this paper further systematically elaborates on the classification of Agent types and relationships. This includes defining three fundamental Agent types: Absolute Zero Agent or Alpha Agent (all five capabilities at zero), Omniscient and Omnipotent Agent or Omega Agent (all five capabilities at infinity), and Finite Agent; it further refines these into 243 theoretically existing Agent types. Concurrently, the paper explores 15 fundamental types of inter-Agent relationships along three dimensions: perception, communication, and interaction.

Generalized Agent Theory not only provides a description of the static structure of Agent but also introduces a refined evolutionary dynamics mechanism to understand the dynamic transformation of Agent capabilities. This mechanism is established upon Agent capability space, with Polar Intelligent Field Model at its core. This model elucidates how the evolution of Agent is influenced by the combined effects of Alpha Degradation Field ( $\vec{F}_\alpha$ ), which points towards capability decay, and Omega Enhancement Field ( $\vec{F}_\Omega$ ), which points towards capability enhancement, with these fields in turn generating corresponding intelligent evolution forces. The resultant of these forces (i. e., Net Intelligent Evolution Force  $\vec{f}_{NIE}$ ) guides Agents migration towards Alpha Polar(Alpha Agent) or Omega Polar(Omega Agent) within capability space. Critically, this dynamics mechanism also introduces a key intrinsic attribute of Agent—Wisdom (W)—to measure the response characteristics of Agent in utilizing it against these evolutionary forces, thereby more comprehensively delineating the evolutionary trajectory of Agent.

The universality of Generalized Agent Theory enables it to deeply analyze some long-standing scientific challenges. In the study of intelligence and consciousness, the theory defines intelligence as the overall efficacy and adaptability exhibited by Agent when utilizing its five core functions to achieve evolutionary goals under the drive of Intelligent field. Consciousness, in turn, is explicitly demarcated as the Control function (Con) itself and its operational process. It is further classified into types such as self-consciousness and other-consciousness based on the origin of control commands (e. g., from internal creation or external input), providing criteria for analyzing the consciousness state of current artificial intelligence, such as large language models.

More groundbreakingly, Generalized Agent Theory attempts to conceptualize the "observer" as Agent, providing a completely new perspective for interpreting several core issues in physics. The theory argues that the Universe itself can be regarded as a dynamically evolving generalized Agent, currently in a finite state. On this basis, objective reality versus subjective non-reality, determinism versus indeterminism, and indeed the nature of time and space, are all interpreted as relative

concepts closely related to the capability state and internal construction processes of Observer Agent, rather than absolute ontological attributes. Crucially, by adjusting the five-dimensional capability vector of Observer Agent (particularly information input, output, and internal processing capabilities), the theory proposes a new pathway for unifying classical mechanics, relativity, and quantum mechanics: the differences among these three theoretical frameworks are rooted in the differing capability configurations of their implicit observer models. For example, the omniscient passive observer of classical mechanics, the information-input-restricted observer of relativity, and the capability-limited, environment-interacting observer of quantum mechanics each construct different models of physical reality. Similarly, the origin of entropy and its observer-dependence are given a new interpretation: entropy increase is considered the inevitable process of increasing information deficit experienced by a limited-capability observer when faced with the fundamental units of the Universe exploring state space under the drive of Omega Enhancement Field. These systematic explorations reveal the immense potential of Generalized Agent Theory as a foundational theoretical tool for connecting different scientific domains and deepening our understanding of the nature of the Universe and intelligence.

Therefore, the core contributions of Generalized Agent Theory, as systematically constructed and elaborated in this paper, are twofold: Firstly, originating from first principles, it establishes a structurally self-consistent and universally explanatory unified theoretical framework. This framework is built upon Standard Agent Model as its core constructive unit, elucidates its evolutionary dynamics through Pole Intelligent Field Model, and institutes a systematic theory of Agent classification and relationships. Secondly, the application of this theory provides insightful solutions to fundamental scientific challenges: it offers novel interpretations and analytical criteria for exploring the nature of intelligence and consciousness; crucially, by conceptualizing the observer in physics as Agent, it opens new research pathways of profound potential for understanding the dynamic evolution of the Universe, the interplay between objective reality and subjective cognition, the origins of determinism and indeterminism, the deep structure of time and space, the nature of entropy, and indeed, for tackling the unification challenge of classical mechanics, relativity, and quantum mechanics. In summary, Generalized Agent Theory aims to provide a unified foundational dialogue platform and research paradigm for key fields such as artificial intelligence, physics, complexity science, and the philosophy of technology, laying a solid groundwork for future theoretical innovations and interdisciplinary integration.

The construction of Generalized Agent Theory has undergone more than a decade of iterative development. It began in 2014, with the initial research goal of seeking a unified standard for measuring generalized intelligent systems (encompassing biological, machine, and artificial intelligence systems), which led us to realize the urgent need for a more fundamental and universal theoretical framework. Drawing inspiration from the von Neumann architecture in computer science, Standard Agent Model was first constructed [5]. Subsequently, in 2017, an intelligence level test scale designed based on this model was used in empirical research [6]. We conducted unified tests on leading AI systems of the time and human subjects of different age groups, revealing stage-specific characteristics of AI intelligence levels and providing a baseline for tracking the significant leap in AI capabilities up to 2024.

The phase of theoretical deepening was marked by the year 2020[7], when we established the two extreme boundary forms of Agent: Alpha Agent, with all capabilities at zero, and Omega Agent, with all capabilities at infinity. From this, we proposed Alpha Gravity and Omega Gravity as the respective driving forces for Agent's evolution towards these two poles. This preliminarily established the dynamic evolutionary mechanism of Generalized Agent Theory, laying a foundation for research into the nature of intelligence and consciousness. By June 2024, by integrating the 'observer' as a special case of Standard Agent Model, we extended Generalized Agent Theory to the analysis of fundamental problems in physics [3], including the nature of reality and non-reality, determinism and indeterminism, and time and space. We also proposed for the first time that the essential differences among classical mechanics, relativity, and quantum mechanics stem from the



intelligence level settings of the implicit observers in each theory. Furthermore, in a study published by our team in March 2025, we elaborated on the view derived from logical deductions based on the two extreme forms of Agent: that the Universe itself is a dynamically evolving Agent, and its fundamental constituent units are also isomorphic Agent [8].

The aforementioned research has laid a forward-looking foundation for the systematic construction and applicational exploration of Generalized Agent Theory. This paper aims to systematically integrate and deepen these research achievements, striving to enhance the rigor, completeness, and scalability of the theory, in order to overcome potential fragmentation issues encountered in earlier explorations.

2. A Unified Structure of Agent Derived from First Principles

The concept of Agent originated in computer science, initially aimed at constructing programs with a degree of autonomy capable of performing tasks on behalf of users. This early notion already encapsulated the fundamental information processing pattern of acquiring input from an environment, performing internal computations and judgments, and generating operational outputs [9].

In artificial intelligence, the concept of Agent has been formalized and extensively studied. Classical definitions, such as that proposed by Russell and Norvig [10], describe Agent as anything that perceives its environment through sensors and acts upon that environment through actuators. The core lies in the perception-action cycle, with internal mechanisms responsible for connecting these two, essentially involving the processing of input information and decision-making. The rise of Distributed Artificial Intelligence (DAI) and Multi-Agent Systems (MAS) further introduced a social dimension to Agent, emphasizing communication and collaboration based on information exchange (e. g., message passing) [11].

In the 21st century, cutting-edge explorations such as embodied intelligence, reinforcement learning, and Agent based on large models have significantly expanded the capability boundaries and application complexity of Agent [12]. However, throughout the evolutionary trajectory of the Agent concept, its fundamental workflow has consistently adhered to an information processing system paradigm. Despite vast differences in specific forms and functional manifestations (including perception, action, internal state maintenance, reasoning, planning, learning, interaction, and even creation), all can fundamentally be attributed to the processes of acquiring, processing, transforming, and utilizing information [13]. A systematic review of the Agent concept clearly reveals its essence as an information processing system operating in a general environment, with its core lying in efficient and flexible information processing capabilities,As shown in Figure 1.

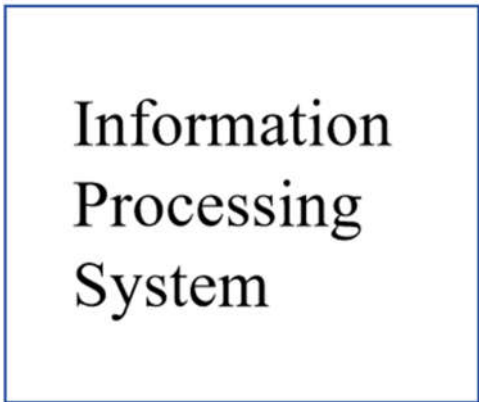
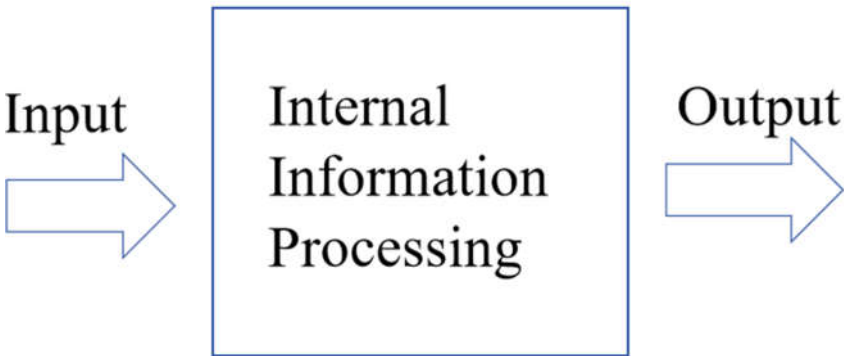


Figure 1. Simplest model of Agent.

This research, adhering to first principles, starts from the most fundamental constituents and operational mechanisms of Agent, aiming to construct a unified theoretical model with robust universal explanatory power. This model strives for 'maximal functional optimality and universal explanatory power' under the premise of 'structural simplicity', while ensuring its abstraction 'granularity' achieves a 'balanced Pole of being appropriate and irreducible'.

The preceding systematic review of the Agent concept clearly reveals that its core essence can be summarized as an information processing system operating in a general environment. Nevertheless, establishing a universal and robust theoretical definition for Agent, particularly one that forms a unified framework capable of guiding its architectural design, remains a fundamental challenge in artificial intelligence. Existing attempts at definition often face a dilemma: on one hand, defining Agent by enumerating overt behavioral characteristics tends to lead to definitions that fragment or even expand indefinitely with technological advancements, making it difficult to form a stable theoretical cornerstone; on the other hand, abstracting the essence of Agent to a singular 'information processing system', while capturing its core attribute and reflecting first principles, is often too generalized and lacks the structured insights necessary for guiding concrete design.

Starting from the fundamental postulate that Agent is an information processing system, we have deeply analyzed the minimal necessary stages of information flow. The operation of Agent can be abstracted as a continuous cycle: acquiring environmental information through perception and transmitting it internally (input); internal algorithms and models processing this information to generate new information or decisions (internal information processing); and finally, producing outputs that affect the environment or interact with other entities through actuators (output). This 'Input-Internal Processing-Output' trimodular model,As shown in Figure 2, while more specific than the definition of 'Agent as an information processing system,' still lacks sufficient descriptive and explanatory power. Particularly, the abstraction granularity of its 'internal information processing' module is too coarse, failing to effectively identify and differentiate more refined information processing functions within Agent—functions that, at a conceptual level, should share an equally important functional status with the input and output modules.



**Figure 2.** Three-module model of Agent.

Such structural limitations are anticipated to pose significant cognitive and modeling obstacles when addressing major theoretical challenges in the future, such as constructing Artificial General Intelligence (AGI), elucidating the origin of consciousness, discerning the deep connections between relativity and quantum mechanics, investigating the physical nature of entropy, and understanding emergent behaviors in complex systems.

To further deepen the understanding of Agent's information processing flow, we examined the von Neumann architecture that guides modern computer design. With its five core components—input, output, storage, computation, and control—this architecture has achieved remarkable success in general-purpose computing. However, when attempting to apply it directly to the high-level, universal modeling of Agent, its inherent limitations become apparent. Re-examining Agent's

information processing logic from first principles, we identified shortcomings in the von Neumann architecture's direct mapping of Agent's core capabilities and, based on this, propose the following two structural adjustments and additions:

First, regarding the 'computation' function: in Agent's information processing flow, computation is essentially the process of operating on stored data elements and established rules to derive new information. Crucially, this newly generated information fundamentally remains subordinate to the generalized set of original information. Therefore, abstracting the computation function and integrating it into the unified concept of Dynamic Storage more accurately reflects computation's intrinsic dependence on storage and their synergy. This move also resonates with the currently prominent 'processing-in-memory' engineering paradigm.

Second, recognizing the Creation function as a key capability distinguishing Agent from classical computational systems holds crucial theoretical significance. It is not merely limited to transforming existing information or logical deduction; its deeper connotation lies in the active generation of entirely new and unanticipated information and solutions. This capability embodies Agent's response to uncertainty, its realization of true autonomy, and its capacity for innovative breakthroughs, and thus warrants consideration as an independent, fundamental functional dimension. The evident inability of classical computational models like the von Neumann architecture to accommodate achievements of human knowledge creation, such as Newton's discovery of universal gravitation or Einstein's formulation of relativity, within their explanatory frameworks, conversely attests to the unique and significant value of the Creation function.

Based on the foregoing first-principles analysis of Agent as an information processing system, and by carefully absorbing the insights and inspiration from the von Neumann architecture, we have identified and refined a novel unified model that adheres to first-principles construction. This model comprises five fundamental functional modules: Information Input, Dynamic Storage, Information Creation, Information Output, and Control. Originating from the minimal necessary stages of information flow, these five functional modules constitute the axiomatic foundation of this unified model of Agent. This model aims to capture the universal and essential attributes of Agent, providing an irreducible theoretical starting Pole for cross-domain research. We name this unified model Standard Agent Model,As shown in Figure 3.

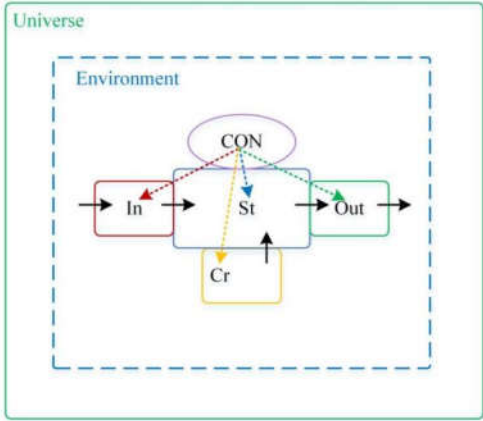


Figure 3. Structure Diagram of Standard Agent Model.

3. Definition and Mathematical Expression of Standard Agent Model

Based on the preceding analysis, this section will provide a detailed definition of the constitution of Standard Agent Model and a mathematical description of its functional modules.

Definition 3. 1: Standard Agent Model (SAM)

Any Agent is composed of five fundamental functional modules: Information Input, Information Output, Dynamic Information Storage, Information Creation, and Information Processing Control。 which can be formalized as:

$$A = \text{Con}(\text{In}, \text{Out}, \text{DS}, \text{Cr})$$

The respective capabilities of these five fundamental functions of Agent constitute the capability of Agent, formalized as:

$$\text{Cap}(A) = (\text{Cap}_{\text{con}}, \text{Cap}_{\text{cr}}, \text{Cap}_{\text{DS}}, \text{Cap}_{\text{out}}, \text{Cap}_{\text{In}})$$

### 3.1. The definitions of each functional module and its corresponding capability

#### 1. Information Input function (In)

This is the functional module through which Agent perceives and acquires information from its external environment E, forming an information input space  $S_{\text{In}}$ . The Information Input function includes, but is not limited to, text input, image input, video input, or any other type of input method. Formally, In is a function from the environmental space ( $S_E$ ) to the agents input space ( $S_{\text{In}}$ ):

$$\text{In}: S_E \rightarrow S_{\text{In}}$$

- $S_E$  represents the space of all information that Agent can, in principle, perceive from its environment E.
- $S_{\text{In}}$  represents the space of raw perceptual data or information representations acquired by Agent from its environment, equivalent to an input information buffer, becoming the initial form accessible for further processing by Agents storage space.

Information input capability, denoted by  $\text{Cap}_{\text{In}}$ , is a measure of the strength of Agents input function. It is specifically quantified by the information-carrying capacity or state-discrimination finesse of the input space  $S_{\text{In}}$  within one input cycle. Therefore, we define capability  $\text{Cap}_{\text{In}}$  as a measure of the scale of  $S_{\text{In}}$ . For simplification, we consider  $S_{\text{In}}$  as a discrete space, and its capacity and resolution are measured by its cardinality (size):

$$\text{cap}(\text{In}) = |S_{\text{In}}|$$

where  $|S_{\text{In}}|$  represents the number of distinct elements in the set  $S_{\text{In}}$ . The range of  $\text{Cap}_{\text{In}}$  is  $0 \leq |\text{Cap}_{\text{In}}| \leq \infty$ . A larger value of  $|S_{\text{In}}|$  indicates that Agent can distinguish and represent more discrete forms of raw input, meaning its capacity and resolution for raw input representation are higher. When  $|S_{\text{In}}| = 0$ , it signifies that Agent lacks information input capability.

#### Information Output function (Out)

This is the functional module through which Agent exerts influence on and modifies the external environment, forming an information output space  $S_{\text{Out}}$ . The Information Output function includes, but is not limited to, image output, text output, action output, and any other type of information output method. It maps internal state information from the dynamic storage space  $S_{\text{DS}}$  to an additive change or command  $a \in S_{\text{Out}}$  upon the environmental state. The information flow must pass through  $S_{\text{DS}}$ . Formally defined as a function:

$$\text{Out}: S_{\text{DS}} \rightarrow S_{\text{Out}}$$

- $S_{\text{DS}}$ : The core internal state space managed by the Dynamic Storage module.



- $S_{Out}$ : The space constituted by all possible additive environmental state changes caused by Agents output. Each  $a \in S_{Out}$  represents a specific change command.
- The environmental state space  $S_E$  must support this additive change, i. e.,  $\forall E \in S_E, \forall a \in S_{Out}, \exists E' \in S_E$  such that  $E' = E + a$ .

Information output capability, denoted by  $Cap_{Out}$ , is a measure of the strength of Agents output function. It is specifically quantified by the capacity and resolution of the environmental state change space ( $S_{Out}$ ) that the Out function can produce in one output cycle. This represents the potential richness and finesse of environmental changes Agent can, in principle, generate in that cycle. For simplification, we consider  $S_{Out}$  as a discrete space, and its capacity and resolution are measured by its cardinality (size):

$$Cap_{Out} = |S_{Out}|$$

where  $|S_{Out}|$  represents the number of distinct elements in the set  $S_{Out}$ . The range of  $Cap_{Out}$  includes all non-negative integers or countable infinity, i. e.,  $0 \leq |Cap_{Out}| \leq \infty$ . A larger value of  $|S_{Out}|$  indicates that Agent can, in principle, produce and distinguish more types of discrete environmental changes within one output cycle, meaning its capacity range and finesse for influencing the environment are higher.

### Dynamic Storage function (DS)

This is the functional module Agent uses for internal dynamic information preservation, responsible for maintaining and managing its dynamic state space  $S_{DS}$ . It includes, but is not limited to, maintaining, integrating, processing, computing, reasoning, deleting or forgetting, or any other type of dynamic storage-related methods. It integrates new information from Information Input  $S_{In}$  and Information Creation  $S_{Cr}$ . A key constraint is that information from  $S_{In}$  and  $S_{Cr}$  must first be incorporated into  $S_{DS}$ . All internal computations, reasoning, etc., operate based on existing information and rules within  $S_{DS}$ , and their results also update and are stored within  $S_{DS}$ ; therefore, computation and reasoning are part of the Dynamic Storage function. This module also covers mechanisms for active deletion or passive forgetting of information.

The Dynamic Storage function achieves state updates through a function DS, which maps the current  $S_{DS}$  state, input information, and created information to the next  $S_{DS}$  state:

$$DS: S_{DS} \times S_{In} \times S_{Cr} \rightarrow S_{DS}$$

- $S_{DS}$ : The information space managed by the Dynamic Storage module.
- $S_{In}$ : The information space of the Information Input function (In).
- $S_{Cr}$ : The information space of the Information Creation function (Cr).

Dynamic storage capability, denoted by  $Cap_{DS}$ , is a measure of the strength of the Dynamic Storage function. It is specifically quantified by the capacity and resolution of the dynamic storage space ( $S_{DS}$ ) it manages. For simplification, we consider  $S_{DS}$  as a discrete space, and its capacity and resolution are measured by its cardinality (size):

$$Cap_{DS} = |S_{DS}|$$

where  $|S_{DS}|$  represents the number of distinct elements in the set  $S_{DS}$ . The range of  $Cap_{DS}$  is  $0 \leq |Cap_{DS}| \leq \infty$ . A larger value of  $|S_{DS}|$  indicates that Agent can store and distinguish more internal states, meaning its capacity and resolution for memory and information preservation are higher.  $|S_{DS}| = 0$  signifies that Agent lacks dynamic storage capability.

### Information Creation function (Cr)

This is the functional module through which Agent can generate information with genuine novelty (new elements or rules not directly deducible from existing information); these outputs form the information creation space  $S_{CR}$ . Operations of the Information Creation function include, but are not limited to: discovering and abstracting new patterns and laws; constructing novel elements or structures through combining, dividing, transforming, or fusing existing information units; and generating entirely new cognitions, insights, or solutions by employing analogical reasoning, forming hypotheses (conjectures), conducting thought experiments, or exploratorily redefining problems, or any other operations that produce genuinely novel information.

Formally, Cr is a function from Agents internal state space ( $S_{DS}$ ) to the information creation space ( $S_{CR}$ ):

$$Cr: S_{DS} \rightarrow S_{CR}$$

- $S_{DS}$  represents the information space managed by the Dynamic Storage module (DS).
- $S_{CR}$  represents the information space formed by the Information Creation function (Cr).

Information creation capability, denoted by  $Cap_{Cr}$ , is a measure of the strength of the Information Creation function. It is specifically quantified by the capacity and resolution of the novel information space ( $S_{CR}$ ) that the Cr function can produce in one cycle. This represents the potential richness and finesse of Agents innovative behavior in that cycle. For simplification, we consider  $S_{CR}$  as a discrete space, and its capacity and resolution are measured by its cardinality (size):

$$Cap_{Cr} = |S_{CR}|$$

where  $|S_{CR}|$  represents the number of distinct elements in the set  $S_{CR}$ . The range of  $cap_{Cr}$  is  $0 \leq |Cap_{Cr}| \leq \infty$ . A larger value of  $|S_{CR}|$  indicates that Agent can, in principle, create and represent more distinct forms of novel information within one input cycle, meaning the potential richness and finesse of its creative output are higher.  $|S_{CR}| = 0$  signifies that Agent lacks information creation capability.

### 2. Control function (Con)

This functional module serves as Agents coordination and management core, responsible for orchestrating functions such as Information Input (In), Output (Out), Dynamic Storage (DS), and Information Creation (Cr). Its core output constitutes meta-commands forming the control command space  $S_{Con}$ . Control responsibilities primarily include: setting priorities for information processing and resource utilization; selecting, initiating, and monitoring information processing procedures (such as input, computation, reasoning, creation, output); and coordinating inter-module operations and resolving conflicts.

Control information in  $S_{Con}$  can originate from: (1) Agents innate or preset instructions; (2) control information acquired via the input function, represented in  $S_{DS}$ , and subsequently formed; (3) control information generated by the creation function, represented in  $S_{DS}$ , and subsequently formed.

Formally, the control function may be conceptualized as a function. This function maps the current Dynamic Storage information ( $S_{DS}$ ) and innate, pre-stored control command ( $S_{Con}$ ) to control command space ( $S_{Con}$ ):

$$Con: S_{DS} \rightarrow S_{Con}$$

- $S_{DS}$ : The information space managed by the Dynamic Storage (DS) module.

- $S_{\text{Con}}$ : The meta-command information space managed by the Control (Con) module.

Control capability, denoted by  $Cap_{\text{Con}}$ , is a measure of the strength of the Con function. It is specifically manifested by quantifying the scale and complexity of the set of control meta-commands  $S_{\text{Con}}$  it can produce. This represents the extent of distinguishable decisions Con can make, the finesse with which it can regulate other modules, and its ability to coordinate complex tasks. For simplification, if  $S_{\text{Con}}$  is considered a discrete space, its capability is measured by its cardinality (size):

$$Cap_{\text{Con}} = |S_{\text{Con}}|$$

where  $|S_{\text{Con}}|$  represents the number of distinct meta-commands in the set  $S_{\text{Con}}$ . The range of  $Cap_{\text{Con}}$  is  $0 \leq |Cap_{\text{Con}}| \leq \infty$ . A larger  $Cap_{\text{Con}}$  indicates stronger decision-making and coordination capabilities of the Con module, allowing for more complex and diverse strategies and behaviors to be executed and managed.  $Cap_{\text{Con}} = 0$  signifies that Agent lacks control capability.

### 3.2. Theoretical Significance of Standard Agent Model

Standard Agent Model, with its five-module structure, establishes a universal and robust framework for the unified description of the multifaceted functionalities of Agent. This model systematically maps key attributes of Agent—such as perception, action, memory and learning, reasoning and decision-making, exploration and innovation, and autonomous adaptation—onto the five core functional modules and their interactions.

The theoretical roots of Standard Agent Model synthesize foundational contributions from fields such as information theory and computation (Shannon[14], Turing[15]), cybernetics (Wiener[16]), the theory of knowledge hierarchy (Ackoff[17], Newell[18]), and computational creativity (Boden[19]). Its core breakthrough lies in its pioneering integration of key cross-disciplinary insights into a functionally complete and internally consistent theoretical system, thereby transcending the limitations of previous fragmented functional analyses.

The more profound theoretical value of Standard Agent Model lies in its potential as an open meta-theory. Through its five fundamental functional modules, it proposes a set of “axiomatic postulates” for the constitution of intelligence, framing it not as a closed, static construct, but as an open structure designed for empirical validation and continuous evolution. This openness is specifically manifested in its inherent modular extensibility: for instance, any novel paradigm of information interaction discovered in the future, such as perception based on quantum states or direct brain-computer interfaces, can be seamlessly integrated as a new implementation of the Information Input (In) module. Likewise, higher-order intelligent behaviors, such as goal emergence driven by self-consciousness, can be absorbed as an extended function of the Information Creation (Cr) module. Therefore, it is precisely this extensibility—rooted in these “axiomatic postulates” and realized through modularity—that ensures the framework remains a robust and effective analytical tool in the face of future scientific breakthroughs, capable of evolving with the times

## 4. Classification and Relationships of Agent Based on Standard Agent Model

In the preceding sections, we derived and defined Standard Agent Model based on first principles. This model comprises five fundamental functional modules, with its corresponding capability vector being:

$$Cap(A) = (Cap_{\text{Con}}, Cap_{\text{Cr}}, Cap_{\text{Ds}}, Cap_{\text{Out}}, Cap_{\text{In}})$$

Standard Agent Model provides us with a unified and foundational framework for analyzing and understanding the essence of Agent. Based on this models 5-dimensional capability vector, we can

systematically classify Agent types and the relationships among multiple Agent instances.

#### 4.1. Classification of Agent

##### 4.1.1. Three Fundamental Classifications of Agent Types

Based on the values of Agents five fundamental capabilities, we can establish two methods for classifying Agent types with different granularities. First, we define three fundamental Agent types that hold conceptual boundary significance:

##### 1. Absolute Zero Agent or Alpha Agent

When all five fundamental capabilities of Agent are zero, i. e. :

$$Cap_{Con} = Cap_{Cr} = Cap_{Ds} = Cap_{Out} = Cap_{In} = 0$$

This Agent is termed Absolute Zero Agent or Alpha Agent, Abbreviated as  $A_\alpha$ . Its capability vector can be concisely represented as:

$$Cap(A) = (0, 0, 0, 0, 0)$$

This type of Agent possesses no information processing capabilities, representing the lower limit of the Agent capability spectrum and a theoretical starting pole. Lifeless systems or objects in physics that are inherently incapable of information processing—such as a stationary stone, an isolated atom, or a propagating photon—can be considered specific instances.

##### 2. Omniscient and Omnipotent Agent or Omega Agent

When all five fundamental capabilities of Agent tend towards infinity, i. e. :

$$Cap_{Con}, Cap_{Cr}, Cap_{Ds}, Cap_{Out}, Cap_{In} \rightarrow \infty$$

This Agent is termed Omniscient and Omnipotent Agent or Omega Agent, Abbreviated as  $A_\Omega$ . Its capability vector can be intuitively represented as:

$$Cap(A_\Omega) = (\infty, \infty, \infty, \infty, \infty)$$

This type of Agent possesses infinite capabilities in information input, output, storage, creation, and control, representing the theoretical upper limit of the Agent capability spectrum.

Discussions of omniscient and omnipotent concepts in some philosophical or theological systems, as well as hypothetical entities in physics with extreme information processing capabilities like Laplaces demon or Maxwells demon, can be seen as its idealized examples.

##### 3. Finite Agent

When the five fundamental capabilities of Agent are neither all zero nor all infinite, i. e., its capability vector:

$$Cap(A_f) = (c_1, c_2, c_3, c_4, c_5)$$

satisfies the condition: there exists at least one capability  $c_i > 0$  and at least one capability  $c_j < \infty$  (where  $i, j \in \{In, Out, DS, Cr, Con\}$ ), this Agent is termed Finite Agent, Abbreviated as  $A_f$ .

Almost all Agent instances encountered in the real world and in current artificial intelligence research belong to this category. Due to the diversity of its capability combinations, Finite Agent cannot be summarized by a single fixed-value vector. Finite Agent constitutes the vast majority of entities in reality and research, a category covering an extremely wide range of objects, including humans, other organisms, computers, robots, and large language models. These three fundamental types provide a first-principles-based fundamental frame of reference for understanding Agent capability space.

##### 4.1.2. Classification of 243 Agent Types

To delineate the diversity of Agent more finely, we can further describe the state of each fundamental capability using a ternary discretization, thereby forming 243 Agent types. Considering the extreme cases of capabilities (zero and infinity) and the intermediate state (finite), we divide each capability  $Cap_i$  into three states:

1. State 0: Indicates the capability is zero ( $Cap_i = 0$ ).
2. State 1: Indicates the capability is finite ( $0 < Cap_i < \infty$ ).
3. State 2: Indicates the capability is infinite ( $Cap_i \rightarrow \infty$ ).

Arranged in the order of capabilities "Control, Creation, Storage, Output, Input," the capability state of Agent can be represented by a five-digit vector. Since each position has three possible states (0, 1, 2), this constitutes a ternary encoding system. Therefore, based on the five fundamental capabilities of Standard Agent Model, there are theoretically  $3^5 = 243$  different Agent types.

These types are indexed starting from 1, with their capability vectors arranged in ascending order of ternary numbers:

- Agent of Index 1: Capability vector (0,0,0,0,0), corresponding to the aforementioned Absolute Zero Agent.

- Agent of Index 2: Capability vector (0,0,0,0,1), representing Agent possessing only finite input capability.

...

- Agent of Index 122: Capability vector (1,1,1,1,1), representing Agent with finite capabilities in input, dynamic storage, creation, output, and control.

- Agent of Index 241: Capability vector (2,2,2,2,0), representing Agent with infinite input, dynamic storage, creation, and control capabilities, but no output capability.

- Agent of Index 242: Capability vector (2,2,2,2,1), representing Agent with infinite input, dynamic storage, creation, and control capabilities, but finite output capability.

- Agent of Index 243: Capability vector (2,2,2,2,2), corresponding to the aforementioned Omniscient and Omnipotent Agent.

For example, a computer hard drive that can store but not read can be considered as possessing only storage and input capabilities (assuming writing is input to the drive itself, and its own output is reading), its capability vector could be expressed as (0,0,1,0,1) (if input is control-related or if it has minimal control for storage; this example directly translates the text, its mapping to "Control, Creation, Storage, Output, Input" order requires full theoretical context).

A typical modern computer system can be approximated as Agent with finite input, output, storage, and control capabilities, but near-zero creation capability, its capability vector representable as (1,0,1,1,1). Humans, as representative organisms, generally possess finite input, output, storage, creation, and control capabilities, their capability vector representable as (1,1,1,1,1).

It is particularly important to note Agent of Index 241, with capability vector (2,2,2,2,0), meaning its output capability is 0 while the other four capabilities are infinite; we name this "**Omniscient Agent**." This type of Agent will play an important role in the analysis of the observer in physics.

For Agent with indices {2, 4, 5, 10, 11, 13, 14, 28, 29, 31, 32, 37, 38, 40, 41, 82, 83, 85, 86, 91, 92, 94, 95, 109, 110, 112, 113, 118, 119, 121, 122}, their five capabilities are neither all 0, nor is any single capability infinite; these 31 Agent types are termed "**Truly Finite Agent**" types.

This classification method based on capability state combinations not only covers extreme boundary-type Agent instances but also provides a systematic framework for describing and



comparing the capability configurations of various Finite Agent instances, helping to lay a foundation for subsequent analysis of fundamental scientific theories based on Standard Agent Model.

#### 4.2. Relationship Types of Agent

After defining the classification system for individual Agent, understanding the interaction patterns among multiple Agent instances is equally crucial. This subsection aims to preliminarily explore the basic types of relationships that can exist between two Agent instances (denoted as A and B), introducing concise mathematical representations and examples to lay the groundwork for subsequent analysis of multi-Agent systems. We propose to characterize inter-Agent relationships from three dimensions: perception, communication, and interaction.

##### 4.2.1. Perception Relationship

Perception Relationship describes the ability and manner in which Agent A and B detect or recognize each others existence. Based on the application of the Information Input (In) capability in Standard Agent Model, the perception state of one Agent towards another can be divided into three categories:

- Unaware: Agent A completely fails to detect the presence of Agent B.
- Indirect Perception: Agent A infers the existence or state of Agent B by observing changes in the shared environment or traces left by B.
- Direct Perception: Agent A can directly receive signals from Agent B or recognize its specific identifiers through its input channels.

Since perception is mutual, As perception state towards B and B's perception state towards A can be independently combined. Therefore, between two Agent instances, there are  $3 \times 3 = 9$  different binary perception relationship states. This perceptual level of connection is a prerequisite for more complex interactions to occur.

Mathematical Representation: We can assign numerical values to these three perception states:  $P = 0$  (Unaware),  $P = 1$  (Indirect Perception),  $P = 2$  (Direct Perception). Then, the binary perception relationship state between A and B can be concisely represented as an ordered pair:

$$R_P(A, B) = (P_{A \rightarrow B}, P_{B \rightarrow A}), \text{ where } P_{A \rightarrow B}, P_{B \rightarrow A} \in \{0, 1, 2\}.$$

Example: A stone (A) and a distant star (B) are typically unaware of each other; their relationship state is (0,0). Humans (A) indirectly perceive dark matter (B) through gravitational effects, while dark matter (B) is generally considered unable to perceive humans (A); this can be represented as (1,0). Two humans (A, B) engaging in face-to-face conversation directly perceive each other; the state is (2,2).

##### 4.2.2. Communication Relationship

Communication Relationship builds upon Perception Relationship, focusing on the possibility and patterns of information exchange between Agent A and B through their respective Information Output (Out) and Information Input (In) capabilities. Assuming both parties can at least indirectly or directly perceive each other, their communication relationship can be divided into four basic types:

- 1.No Mutual Communication: Neither A nor B can effectively receive information output by the other.
- 2.Bidirectional Communication: Both A and B can effectively output information to each other and receive and process each others information through their input channels.
- 3.Unidirectional Communication  $A \rightarrow B$ : A can output information to B, and B can effectively receive it, but the reverse is not true.

4. Unidirectional Communication  $B \rightarrow A$ : B can output information to A, and A can effectively receive it, but the reverse is not true. This relationship clarifies the basic structure of information flow between Agent instances, forming the basis for advanced interactions such as collaboration or conflict.

Mathematical Representation: Let the Boolean variable  $C_{X \rightarrow Y}$  indicate whether an effective communication channel exists from Agent X to Agent Y ( $C_{X \rightarrow Y} = 1$  indicates existence,  $C_{X \rightarrow Y} = 0$  indicates non-existence). Then the communication relationship state between A and B can be represented by an ordered pair:

$$R_C(A, B) = (C_{A \rightarrow B}, C_{B \rightarrow A})$$

The four types correspond to: No Mutual Communication (0,0); Bidirectional Communication (1,1); Unidirectional  $A \rightarrow B$  (1,0); Unidirectional  $B \rightarrow A$  (0,1).

Example: A dog (A) barks at a stone (B); no effective communication is formed, state (0,0). Two people (A, B) talk using walkie-talkies; this is bidirectional communication (1,1). A television station (A) sends signals to a television set (B); this is unidirectional  $A \rightarrow B$  (1,0). An environmental sensor (B) sends readings to a data center (A), while the data center does not send commands to that sensor; this is unidirectional  $B \rightarrow A$  (0,1).

#### 4.2.3. Interaction Relationship

Interaction Relationship further describes the dynamic interaction patterns formed between Agent A and B based on their goals, intentions, and control capabilities, given their perceptual and communication abilities. This relationship profoundly affects the behavior and evolution of multi-Agent systems. Based on the nature of the interaction, several basic types can be distinguished:

1. Neutral / Non-Cooperative: The behaviors of both parties are independent and do not interfere with each other.
2. Equal Cooperation: Both parties collaborate as equals, pursuing shared or compatible goals.
3. Adversarial / Competitive: The goals of both parties conflict; they oppose each other or compete for resources.
4. Hierarchical Control: An asymmetric influence or dominance relationship exists, including two modes: A controls B ( $A \rightarrow B$ ) and B controls A ( $B \rightarrow A$ ). This set of interaction relationship classifications, combined with the aforementioned perception and communication relationships, provides more refined conceptual tools for in-depth understanding and modeling of the complex dynamics of multi-Agent systems.

Mathematical Representation: We can assign discrete identifiers to these basic interaction modes. For example, define the set of interaction relationship types  $T_I$ :

$$T_I = \{\text{Neu}, \text{Equ}, \text{Adv}, \text{HC}_{A \rightarrow B}, \text{HC}_{B \rightarrow A}\}$$

representing Neutral, Equal Cooperation, Adversarial, A controls B, and B controls A, respectively. Then, the specific interaction relationship between A and B can be represented by a state variable:

$$R_I(A, B) \in T_I$$

Selecting the corresponding symbol from  $T_I$  indicates the current interaction relationship mode.

Example: The relationship between two unfamiliar cars (A, B) driving according to rules on a highway is typically Neutral (Neu). Two companies (A, B) forming a joint venture to cooperate equally have an Equal Cooperation relationship (Equ). Soldiers from two opposing countries (A, B) engaging in battle on a field are in an Adversarial relationship (Adv). In a company, a manager (A) assigning tasks to a subordinate (B) reflects a Hierarchical Control relationship ( $\text{HC}_{A \rightarrow B}$ ).

## 5. The Evolutionary Dynamics of Agent: A Mechanism Based on Standard Agent Model

Standard Agent Model provides us with a static framework for describing the core capabilities of Agent, upon which classifications of types and relationships have been made. However, a key characteristic of Agent is its dynamism—its capabilities are not immutable but continuously evolve through interaction with the environment and internal operations. Understanding the intrinsic laws and driving forces behind these capability state transformations is a crucial component of Generalized Agent Theory. Within the framework of Standard Agent Model, we will introduce a field-based perspective to explore the dynamics mechanism of Agent capability state transitions, striving for independence from any a priori temporal background and focusing on the structure of capability space itself and its intrinsic driving forces.

### 5.1. Agent Capability Space

The capability state of Agent A is defined by its capability vector  $Cap(A)$ :

$$Cap(A) = \mathbf{c} = (Cap(A)_{con}, Cap(A)_{cr}, Cap(A)_{ds}, Cap(A)_{out}, Cap(A)_{in})$$

All such capability vectors  $\mathbf{c}$  constitute the high-dimensional Agent Capability Space ( $S_{Cap}$ ).

Within this space, two Poles exist as theoretical references, corresponding precisely to the two boundary Agent types we defined in previous sections based on Standard Agent Model:

#### Definition 5.1 Alpha Pole ( $A_\alpha$ ):

This pole corresponds to Alpha Agent, whose capability vector is the zero vector  $\mathbf{c} = \vec{0}$ , representing the absolute zero or baseline state of capabilities.

#### Definition 5.2 Omega Pole ( $A_\Omega$ ):

This pole corresponds to Omega Agent, whose capability vector is the limit state where all components are infinite,  $\mathbf{c} \rightarrow \vec{\infty}$ , representing the idealized state of infinite capabilities.

Any Finite Agent ( $A_f$ ) occupies a state pole  $\mathbf{c}_f$  within Agent Capability Space  $S_{Cap}$  between  $A_\alpha$  and  $A_\Omega$ . The "evolution" of Agent is the migration of its state pole within  $S_{Cap}$ . To understand the driving forces behind this migration, we propose Pole Intelligent Field Model. This model posits that Agent Capability Space  $S_{Cap}$  is permeated by two fundamental, opposing "fields of force" that collectively determine the tendency of Agent capability transformation.

### 5.2. Pole Intelligent Field Model

To understand the driving forces behind Agents migration in Capability Space, we propose Pole Intelligent Field Model. This model posits that Agent Capability Space  $S_{Cap}$  is permeated by two fundamental, opposing "fields of force" that collectively determine the tendency of Agent capability transformation, As shown in Figure 4.

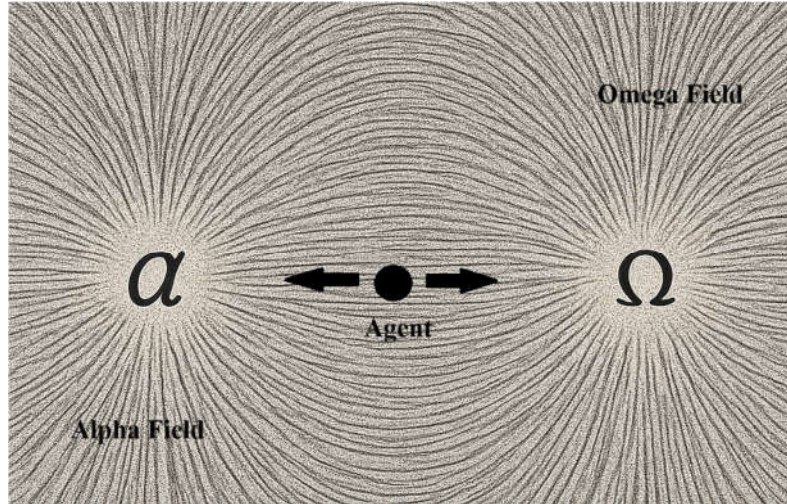


Figure 4. Diagram of Intelligent Pole Intelligent Field Model.

**Definition 5. 3 Alpha Degradation Field (Alpha Field,  $\vec{F}_\alpha$ ):**

This is a vector field defined throughout Agent Capability Space  $S_{\text{Cap}}$ , referred to as Alpha Field and denoted as  $\vec{F}_\alpha(c, E)$ . It depends on the current capability state  $c$  of Agent and its environment  $E$ . It represents the aggregate effect of all factors causing capability decay, simplification, or regression towards the baseline state. These factors may stem from resource consumption, information forgetting, structural degradation, environmental noise interference, and the intrinsic costs of maintaining capabilities. The direction of  $\vec{F}_\alpha$  generally points towards diminishing capabilities, i. e., tending to "pull" the state Pole  $c$  towards Alpha Pole  $A_\alpha$  ( $c = 0$ ).

**Definition 5. 4 Omega Enhancement Field (Omega Field,  $\vec{F}_\Omega$ ):**

This is likewise a vector field defined on  $S_{\text{Cap}}$ , referred to as Omega Field and denoted as  $\vec{F}_\Omega(c, E)$ , also dependent on state  $c$  and environment  $E$ . It represents the aggregate effect of all factors promoting capability enhancement, complexification, or development away from the baseline state. These factors are associated with processes such as learning, adaptation, creation ( $\text{Cap}_{\text{Cr}}$ ), effective resource acquisition and utilization, self-organization, and synergistic effects. The direction of  $\vec{F}_\Omega$  generally points towards enhancing capabilities, i. e., tending to "push" the state Pole  $c$  away from Alpha Pole  $A_\alpha$  and towards Omega Pole  $A_\Omega$  ( $c \rightarrow \infty$ ) within capability space.

To more precisely describe the evolutionary impact of these fields on a specific Agent, we hereby introduce the concept of "force," reifying the effect of the field on a specific Agent.

**Definition 5. 5 Alpha Force ( $\vec{f}_\alpha$ )**

Alpha Force is the specific evolutionary driving force exerted by Alpha Degradation Field on Agent A, which is in a particular capability state  $c$  and environment  $E$ , causing its capabilities to tend towards decay, simplification, or regression. For Agent A, the Alpha Force  $\vec{f}_\alpha(A)$  it experiences is equal to the field vector of Alpha Degradation Field at its capability state point:

$$\vec{f}_\alpha(A) = \vec{F}_\alpha(c_A, E_A)$$

**Definition 5.6 Omega Force ( $\vec{f}_\Omega$ )**

Omega Force is the specific evolutionary driving force exerted by Omega Enhancement Field on Agent A, which is in a particular capability state  $c$  and environment  $E$ , prompting its capabilities to tend towards enhancement, complexification, or development. For Agent A, the Omega Force  $\vec{f}_\Omega(A)$  it experiences is equal to the field vector of Omega Enhancement Field at its capability state point:

$$\vec{f}_\Omega(A) = \vec{F}_\Omega(c_A, E_A)$$

*5.3. Net Intelligent Evolution Force and Net Intelligent Evolution Field*

Within Agent capability space  $S_{cap}$ , the diffused Alpha Degradation Field ( $\vec{F}_\alpha$ ) and Omega Enhancement Field ( $\vec{F}_\Omega$ ) interact, collectively shaping the dynamic landscape of Agent capability evolution. At any given capability state point, the ultimate evolutionary drive experienced by Agent is embodied as Net Intelligent Evolution Force, and the spatial distribution of these forces constitutes Net Intelligent Evolution Field.

**Definition 5.7 Net Intelligent Evolution Force ( $\vec{f}_{NIE}$ )**

Net Intelligent Evolution Force is the vector difference between Omega Force and Alpha Force acting upon Agent A, determining the instantaneous final direction and magnitude of its evolution in capability space. The mathematical description of Net Intelligent Evolution Force is:

$$\vec{f}_{NIE}(A) = \vec{f}_\Omega(A) - \vec{f}_\alpha(A)$$

**Definition 5.8 Net Intelligent Evolution Field ( $\vec{F}_{NIE}$ )**

Net Intelligent Evolution Field is the vector distribution of the net evolutionary tendency at every point in capability space, where its field vector at a point is equal to the vector



difference between Omega Enhancement Field and Alpha Degradation Field. Net Intelligent Evolution Force is the manifestation of Net Intelligent Evolution Field on a specific Agent. The mathematical description of Net Intelligent Evolution Field is:

$$\vec{F}_{NIE}(c, E) = \vec{F}_{\Omega}(c, E) - \vec{F}_{\alpha}(c, E)$$

Thus, the relationship between Net Intelligent Evolution Force and Net Intelligent Evolution Field is:

$$\vec{f}_{NIE}(A) = \vec{F}_{NIE}(c_A, E_A)$$

Net Intelligent Evolution Field  $\vec{F}_{NIE}(c, E)$  assigns a vector to every point in capability space, collectively depicting the local dynamic landscape of capability state transitions:

- The **direction** of the field vector indicates the instantaneous direction in which Agent capability is most inclined to evolve at that state  $c$ .
- The **magnitude** (modulus) of the field vector,  $||\vec{F}_{NIE}||$ , reflects the relative instability of that state, or the intensity of the capability transformation tendency. The greater the modulus, the stronger the evolutionary driving force.

**Core Principle of Dynamics:** The state migration of Agent in capability space  $S_{cap}$  (i. e., the evolution of its capabilities) is directly driven by Net Intelligent Evolution Force ( $\vec{f}_{NIE}$ ) generated by Net Intelligent Evolution Field (the vector difference of Omega Enhancement Field and Alpha Degradation Field) acting upon it. The Agents state point tends to shift along the field lines or integral curves of the field corresponding to this force.

Based on the structure of Net Intelligent Evolution Field, we can analyze the fundamental evolutionary trends of Agent in capability space:

### 1. Regions Tending Towards Omega Pole ( $A_{\Omega}$ ):

In these regions of capability space, the influence of Omega Enhancement Field  $\vec{F}_{\Omega}$  dominates Alpha Degradation Field  $\vec{F}_{\alpha}$ , causing the Net Intelligent Evolution Field vector  $\vec{F}_{NIE}$  to generally Pole towards comprehensive capability enhancement, driving Agent to evolve towards  $A_{\Omega}$ .

### 2. Regions Tending Towards Alpha Pole ( $A_{\alpha}$ ):

In these regions, the influence of  $\vec{F}_{\alpha}$  is predominant, causing  $\vec{F}_{NIE}$  to generally Pole towards comprehensive capability decay, driving Agent to regress or degenerate towards  $A_{\alpha}$ .

### 3. Equilibrium/Steady States ( $c^*$ ):

There may exist certain state points or regions in capability space where Net Intelligent Evolution Field is a zero vector, i. e.,  $\vec{F}_{NIE}(c^*, E) = 0$ . This means that at such states, the driving forces for enhancement and decay achieve a dynamic equilibrium, and the system has no net evolutionary

tendency. These equilibrium points can be further distinguished based on the structure of the nearby field:

- **Stable Equilibrium Points (Attractors):** Nearby field lines converge towards these points; Agent tends to return to such a state after minor perturbations.

- **Unstable Equilibrium Points (Repellers or Saddle Points):** Nearby field lines diverge from these points; once Agent deviates, it will continue to move away from such a state. The existence and stability of equilibrium states determine whether Agents capability structure tends to solidify at a specific level or will continuously evolve.

It must be particularly emphasized that Net Intelligent Evolution Field,  $\delta_c(c, E)$ , itself describes the local, time-parameter-independent tendency of transformation for each state Pole  $c$  in capability space. Although the fields structure implies the possibility of evolutionary paths, this does not mean the model a priori sets a unified external time parameter  $t$ . A specific time parameter or evolutionary order parameter  $\tau$  can be introduced upon this base field, for example, by solving the fields integral curves (formalized as  $dc/d\tau = \delta_c(c, E)$ ) to obtain specific evolutionary trajectories, or by modeling it as a Markov process based on state transition probabilities determined by  $\delta_c(c, E)$ . The core of this section is to define this Net Intelligent Evolution Field itself as the basis for capability evolution, independent of specific path parameters, which helps maintain the theorys potential for background independence.

The dynamic mechanism of this model finds broad corroboration in the real world. For instance, in **biological evolution**, species develop more complex adaptive capabilities and tend towards higher capability states through genetic variation (associated with  $Cap_{Cr}$ ) and natural selection (driven by environmental feedback, Omega Field)[20]; simultaneously, environmental pressures, genetic drift, and senescence (Alpha Field) can lead to simplification or extinction, tending towards  $A_a$  [21]. **Individual learning** processes also reflect these dynamics: skills are acquired (enhancing  $Cap_{DS}$ ,  $Cap_{Con}$ , etc.) [22]; through practice and information input (driven by Omega Field), while forgetting and skill decay (reflecting Alpha Field) are ubiquitous, potentially leading to a stable skill level (equilibrium state  $\delta_c = \vec{0}$ ) [23]. In **technological development**, artificial intelligence models (such as large language models) experience rapid capability enhancement (tending towards  $A_a$ ) driven by data and computational power (Omega Field) [24], yet are still limited by architecture, costs, and ethics (constituting Alpha Field); their development is not infinite [25]. These examples demonstrate that the interplay of Alpha Field and Omega Field provides a unified perspective for understanding the evolutionary trends of complex systems across different domains.

#### 5.4. Wisdom (W): The Intrinsic Metric of Agent Evolution and Its Dynamics Equation

In the aforementioned dynamics mechanism, the response of Agent to Net Intelligent Evolution Force is not entirely uniform. To explain the different evolutionary rates exhibited by different Agents towards one of the two intelligent poles under the same intelligent evolution force, we introduce an entirely new fundamental dimension termed Wisdom (W), with W as its dimensional symbol.

##### Definition 5.9 Wisdom (W)

Wisdom (W) is a key intrinsic metric of Agent, a scale measuring the magnitude of its comprehensive information processing capability, used to quantify the total amount of intelligence possessed by Agent. It further reflects its response characteristics when changing its capability state under the influence of intelligent field or intelligent force in capability space, which in some cases manifests as "evolutionary inertia" (e. g., resisting decay), and in others as accelerated utilization of favorable conditions.

The constitution of Wisdom (W) is conjointly determined by the current synergistic structure of the five core functions defined by Standard Agent Model (Input, Output, Dynamic Storage, Creation, Control) and the total amount of information within its dynamic storage. In this paper, we preliminarily define Wisdom (W) as a quantitative measure of the comprehensive capability of Agent and emphatically reveal its critical role in evolutionary dynamics. In future research, we plan to further explore the multidimensional constitution of Wisdom (W); for instance, by systematically distinguishing between "fundamental information processing capabilities," represented by functions such as Input, Output, Dynamic Storage, and Creation, and "advanced regulatory wisdom," centered around the Control function, and by conducting more refined quantification and modeling of both aspects and their interactions.

Unlike mass in classical physics, which is usually a constant scalar, Wisdom (W) is a dynamic variable. It not only reflects the current "accumulation" of Agent but also influences its future "developmental potential," exhibiting the following key characteristics:

1. **Matthew Effect:** If Omega Pole is considered the positive direction of evolution, when Net Intelligent Evolution Force  $\vec{f}_{NIE} > 0$  (force direction points towards Omega Pole), Agent with greater Wisdom (W) can more effectively utilize this force due to its internal structure and information reserves, manifesting as greater evolutionary gain towards Omega Pole, thus achieving faster positive acceleration.

2. **Resilience Effect:** When Net Intelligent Evolution Force  $\vec{f}_{NIE}$  is negative (force direction points towards Alpha Pole), Agent with greater Wisdom (W) can better resist negative impacts due to its structural stability and information redundancy, manifesting as greater evolutionary hindrance towards Alpha Pole, thereby slowing the rate of decay. Thus, Wisdom (W) is not a simple inertial mass but a state-dependent, direction-sensitive response-regulating factor.

The Wisdom Dynamics Equation is expressed as follows:

1. When  $\vec{f}_{NIE}(A) \geq 0$  (Matthew Effect is dominant): where  $k_p$  is a positive proportionality constant

$$\vec{a}_A = \vec{f}_{NIE}(A) \cdot W_A \cdot k_p \quad (\text{where } k_p \text{ is a positive proportionality constant})$$

2. When  $\vec{f}_{NIE}(A) < 0$  (Resilience Effect is dominant): where  $k_n$  is a negative proportionality constant

$$\vec{a}_A = \frac{\vec{f}_{NIE}(A)}{W_A \cdot k_n} \quad (\text{where } k_n \text{ is a negative proportionality constant})$$

This set of equations quantitatively reveals how the external evolutionary pressure (force) acting on Agent, through its dynamic intrinsic attribute (Wisdom), ultimately determines the specific trajectory (acceleration) of its capability evolution. It is important to note that since evolutionary acceleration will lead to changes in the capability state, and Wisdom itself is determined by the capability state (functional structure and total information), the change in Wisdom is a natural

component of the evolutionary process. This implies that Wisdom and capability state will influence each other and co-evolve, reflecting the non-linear and adaptive characteristics of Agent evolution. In specific analysis, this is typically handled through iterative methods of discrete time steps or by quasi-static approximation within a specific timescale.

This framework lays the foundation for a deeper understanding and quantification of the evolutionary differences among various Agents. Wisdom is a core, fundamental parameter or metric defined within Generalized Agent Theory. Given its uniqueness and foundational role in intelligent phenomena, it has the potential to become a new dimension in a broader sense. However, whether it can ultimately be universally accepted by the scientific community as a new fundamental dimension(metric) to existing physical dimensions will depend on more extensive future theoretical verification, empirical research, and compatibility with other disciplines. This exploration still requires continuous in-depth study and observation.

5.5. System Construction of Generalized Agent Theory

In summary, Standard Agent Model constructed from first principles, the Agent classification system and relationship types derived from this model, and the Agent evolutionary dynamics mechanism expounded in this chapter—which is based on Agent capability space, describes evolutionary drive via Polar Intelligent Field Model, and introduces Wisdom (W) as a key intrinsic attribute of Agent and the core object of field interaction—collectively constitute the core theoretical framework of Generalized Agent Theory . This framework provides a unified, foundational conceptual and mathematical toolkit for understanding and analyzing scientific phenomena across different domains and levels.

6. Analysis of Intelligence and Consciousness Based on Generalized Agent Theory

The nature of intelligence and consciousness represents not only core challenges in modern science and philosophy but also key issues in artificial intelligence [26]. Despite significant progress in related research, existing theories (such as Integrated Information Theory, Global Workspace Theory, etc. [27,28]) and definitions are numerous and still face challenges in terms of unity, fundamentality, and addressing the hard problem of subjective experience, highlighting the necessity of constructing a more universal theoretical framework [29].

Generalized Agent Theory, as a unified model of Agent constructed from first principles, offers a unique perspective on this. This chapter aims to utilize the core principles of Generalized Agent Theory—including Standard Agent Model, Agent Capability Space ( $S_{Cap}$ ) and its evolutionary dynamics ( $\alpha$  Pole ( $A_\alpha$ ),  $\Omega$  Pole ( $A_\Omega$ ), Alpha Degradation Field ( $\vec{F}_\alpha$ ), Omega Enhancement Field ( $\vec{F}_\Omega$ ))—to elucidate its definitions of intelligence and consciousness and their interrelationship, rather than conducting an exhaustive comparative analysis of numerous existing theories.

Generalized Agent Theory posits that understanding intelligence and consciousness first requires clarifying the basis of their existence, objectives, driving forces, and intrinsic connections. In this regard, the core viewpoints provided by Generalized Agent Theory are:

**1. Subject Affiliation:** Intelligence and consciousness are inherent attributes of Agent (as defined by Standard Agent Model); their emergence and manifestation are inseparable from Agents structural and functional basis.

**2. Fundamental Objective:** The fundamental role of intelligence and consciousness is to serve Agents survival and evolution within Agent Capability Space  $S_{Cap}$ . Agents specific evolutionary direction (towards  $A_{\Omega}$  or  $A_{\alpha}$ , or maintaining a certain equilibrium state) is not predetermined but depends on the direction of the resultant force from Alpha Degradation Field and Omega Enhancement Field that Agent faces in a specific environment, i. e., it is determined by Net Intelligent Evolution Field.

**3. Intrinsic Driving Force:** The fundamental driving force for the emergence and operation of intelligence and consciousness originates from Pole Intelligent Field Model defined in Generalized Agent Theory, specifically manifesting as the driving force exerted by Net Intelligent Evolution Field formed by the interaction of Alpha Degradation Field ( $\vec{F}_{\alpha}$ ) and Omega Enhancement Field ( $\vec{F}_{\Omega}$ ).

**4. Intrinsic Relationship:** Intelligence is the overall efficacy and adaptability exhibited by Agent in utilizing all its five core functions (Input, Output, Dynamic Storage, Creation, and Control) to evolve towards its fundamental objective under a given driving force. Consciousness, its essence is explicitly demarcated by Generalized Agent Theory as the Control function (Con) and its operational process. Therefore, consciousness is a key component of overall intelligence, rather than a concept parallel to or separate from it.

The subsequent content of this chapter will revolve around these core viewpoints, conducting in-depth discussions on the definitions, assessment methods, type classifications, and interrelationships of intelligence and consciousness according to Generalized Agent Theory.

#### 6.1. Definition and Assessment Methods for Intelligence

Based on Generalized Agent Theory and adhering to first principles, this section will define "intelligence" from three dimensions: its **subject** (the framework of Standard Agent Model), **fundamental objective** (evolutionary direction in capability space), and **intrinsic driving force** (driven by Pole Intelligent Field Model).

##### Definition 6. 1 Intelligence

Within the framework of Generalized Agent Theory, intelligence is defined as: the **overall information processing efficacy and environmental adaptability** exhibited by an Agent conforming to Standard Agent Model, as it utilizes all its five core functions (Information Input (In), Information Output (Out), Dynamic Storage (DS), Information Creation (Cr), and Control function (Con)) under the combined influence of its inherent Alpha Degradation Field ( $F_{\alpha}$ ) and Omega Enhancement Field ( $F_{\Omega}$ ), to adapt to the environment and achieve its evolutionary objective (tending towards  $A_{\alpha}$  ( $\alpha$  Pole) ,  $A_{\Omega}$  ( $\Omega$  Pole) , or a specific equilibrium state) determined by Net Intelligent Evolution Field.

The magnitude or strength of intelligence can be assessed through some comprehensive measure of Agents capability vector  $CapA$ . This capability vector is composed of the capability components of Agents five core functional modules:

$$Cap(A) = (Cap(A)_{con}, Cap(A)_{cr}, Cap(A)_{ds}, Cap(A)_{out}, Cap(A)_{in})$$

There are multiple methods for a comprehensive measure of  $CapA$ . A commonly used method is the **Weighted Sum Method**, which assigns different weights  $w_i$  (where  $\sum w_i = 1, w_i \geq 0$ ) to the five core



capabilities. The comprehensive assessment value of intelligence,  $I_A$ , can then be expressed as:

$$I_A = w_{In} \cdot Cap_{In} + w_{Out} \cdot Cap_{Out} + w_{DS} \cdot Cap_{DS} + w_{Cr} \cdot Cap_{Cr} + w_{Con} \cdot Cap_{Con}$$

The specific values of the weights can be adjusted according to Agents environment, task requirements, or assessment focus.  $Cap_{Con}$ , as the core for coordinating all capabilities, its weight is particularly important when assessing advanced intelligent behaviors. To illustrate the application of the weighted sum method, we set a hypothetical group of weights. Considering the central coordinating role of the Control function (Con) and the importance of the Creation function (Cr) for advanced intelligence, we can set the following weights (this set of weights is merely illustrative; in practical applications, they can be determined by methods such as the Delphi method):

$w_{In} = 0.15, w_{Out} = 0.15, w_{DS} = 0.20, w_{Cr} = 0.25, w_{Con} = 0.25$  (Sum of weights:  $0.15 + 0.15 + 0.20 + 0.25 + 0.25 = 1.00$ )

Now, let us assume capability scores for three different types of Agent (assuming a scoring range of 0-100, where higher scores represent stronger capabilities; these scores are also illustrative), as shown in Table 1.

**Table 1.** Illustrative Capability Level Scores for Three Agent Types.

Agent Type	$Cap_{In}$ ( $\times 0.15$ )	$Cap_{Out}$ ( $\times 0.15$ )	$Cap_{DS}$ ( $\times 0.15$ )	$Cap_{Cr}$ ( $\times 0.15$ )	$Cap_{Con}$ ( $\times 0.15$ )	Intelligence Level ( $I_A$ )
Calculator	$20 \times 0.15$	$10 \times 0.15$	$5 \times 0.20$	$0 \times 0.25$	$5 \times 0.25$	6.75
Intelligent Car	$80 \times 0.15$	$70 \times 0.15$	$75 \times 0.20$	$30 \times 0.25$	$80 \times 0.25$	66.25
Adult Human	$70 \times 0.15$	$75 \times 0.15$	$80 \times 0.20$	$85 \times 0.25$	$90 \times 0.25$	81.25

The weight distribution and capability scores in the example above are highly simplified and subjectively set, merely to demonstrate the calculation process. In actual intelligence assessment, how to objectively and accurately quantify the five core capabilities ( $Cap_{In}$  to  $Cap_{Con}$ ), and how to set reasonable weights  $w_i$  according to specific contexts and assessment objectives, require in-depth research to form executable schemes based on specific situations. For example, some tasks may place more emphasis on creativity ( $w_{Cr}$  being higher), while others may value the stability and precision of control ( $w_{Con}$  being higher). Furthermore, these capabilities may not be entirely independent, and their interactions could also affect overall intelligent performance.

The intelligence assessment framework based on the weighted sum method proposed in this section not only provides a foundational structure for measuring the current intelligence level of Agent, but its calculated comprehensive capability value also holds potential as a preliminary or approximate quantification method for Wisdom (W) as introduced in Chapter 5. This is because the core of Wisdom lies in measuring the comprehensive information processing capability of Agent, and the weighted sum method is precisely a specific calculation of this comprehensive capability.

However, it should also be noted that Wisdom, as defined in Section 5. 4 of this paper, is further emphasized as being conjointly determined by the "current synergistic structure" of the five core functions and the "total amount of information in dynamic storage," and is manifested through unique response characteristics in evolutionary dynamics. Therefore, although the weighted sum method offers a valuable starting point for quantification, future research should be dedicated to exploring more comprehensive and precise methods for measuring the rich connotations of Wisdom.

For instance, considerations should include how to more directly incorporate factors such as the synergistic structure among modules, and the total amount and quality of information in storage, into the computational model of Wisdom, thereby more profoundly revealing its essence as a key intrinsic attribute in the evolution of Agent.

## 6.2. Definition and Classification of Consciousness

Within Standard Agent Model—the core of Generalized Agent Theory—the Control function occupies a unique and critical position. Unlike the other four fundamental functions that directly process information, the core responsibility of the Control function is the overall coordination, management, and regulation of these four modules and the information flows they carry. This characteristic of unified command and management over internal states and external behaviors makes the Control function a natural focal Pole for exploring Agent autonomy, decision-making, and even deeper cognitive phenomena. Although the direct association of consciousness with the control function is not currently mainstream, it has been reflected in multidisciplinary theories[30]. Most existing research views control as a result or condition of consciousness, rather than viewing consciousness itself as a control mechanism[31]. Generalized Agent Theory goes further, directly defining the Control function itself and its operational process as the essence of consciousness, formulated as follows:

### Definition 6.2 Consciousness

Within the framework of Generalized Agent Theory, consciousness is defined as: Agents Control function (Con), by virtue of its unified command capability, conducting overall coordination and advanced regulation of the four core functions—Information Input (In), Information Output (Out), Information Dynamic Storage (DS), and Information Creation (Cr)—under the evolutionary drive formed by Alpha Degradation Field and Omega Enhancement Field, to achieve its fundamental evolutionary objective of tending towards  $\alpha$  Pole or  $\Omega$  Pole in Agent Capability Space ( $S_{Cap}$ ).

As previously elaborated for the Control function (Con), the meta-commands it generates (constituting the control command space  $S_{Con}$ ) can originate from various informational bases: Agents inherent preset instructions, original content generated via the Information Creation function (Cr), and information acquired from the external environment via the Information Input function (In). All this information must be represented and integrated within the Dynamic Storage space ( $S_{DS}$ ) to serve as the basis for the Control functions decision-making.

Based on the meta-commands generated by the Control function Con (whose decision basis stems from different primary sources of information in  $S_{DS}$ ) and the state of Control function capability  $Cap_{Con}$ , consciousness can be distinguished into the following four basic types:

#### 1. Self-Consciousness ( $C_S$ )

- **Core Definition:** Agents control space meta-commands are not empty, or they originate from inherent control meta-commands within the control space, or they originate from control meta-commands generated by its own Information Creation function (Cr) and integrated into  $S_{DS}$ .

- **Characteristics:** Reflects Agent controlling its own information processing capabilities (information input, output, dynamic storage, and creation) by following its own control commands.

- **Mathematical Criterion:**

$\text{Cap}_{\text{Con}} > 0 \wedge (\text{meta-control command decisions primarily originate from } Cr \rightarrow S_{DS} \text{ or originate from } S_{\text{Con}} \text{ itself}). \text{Equivalent to } \text{Cap}(C_S) > 0.$

## 2. Other-Consciousness ( $C_O$ )

- **Core Definition:** Agents control space meta-commands are not empty and originate from control meta-commands acquired by Agent through Information Input and integrated into Dynamic Storage space  $S_{DS}$ .
- **Characteristics:** Reflects Agent essentially controlling its own information processing capabilities (information input, output, dynamic storage, and creation) by following control commands from other Agent instances.
- **Mathematical Criterion:**  
 $\text{Cap}_{\text{Con}} > 0 \wedge (\text{meta-control commands primarily originate from } In \rightarrow S_{DS}). \text{Equivalent to } \text{Cap}(C_O) > 0.$

## 3. Mixed Consciousness ( $C_M$ )

- **Core Definition:** Agent simultaneously possesses self-consciousness and other-consciousness.
- **Characteristics:** Reflects Agent following control commands that include its own as well as those from other Agent instances, operating its own information processing capabilities (information input, output, dynamic storage, and creation).
- **Mathematical Criterion:**  $\text{Cap}(C_S) > 0 \wedge \text{Cap}(C_O) > 0.$

## 4. Unconsciousness ( $C_U$ )

- **Core Definition:** The capability  $\text{Cap}_{\text{Con}}$  of Agents Control function (Con) is equal to zero, meaning its control function is completely missing or ineffective, unable to read Agents control space meta-commands. Or Agents control space meta-commands are empty.
- **Characteristics:** Agent does not possess control function, or lacks effective commands for the control function to call upon.
- **Mathematical Criterion:**  $\text{Cap}_{\text{Con}} > 0$  or  $S_{\text{Con}} = \emptyset.$

Based on the classification of consciousness types by Generalized Agent Theory, a preliminary determination of the consciousness state of current artificial intelligence (AI), particularly large language models (LLMs), can be made. Although LLMs exhibit powerful information processing and seemingly novel text generation capabilities, their operational essence still lies within the Turing computation framework, involving complex computations and information recombination based on preset rules and external data (which constitute their core internal state space  $S_{DS}$ ).

Generalized Agent Theory distinguishes between the Dynamic Storage function (DS), primarily responsible for information preservation and inference, and the Information Creation function (Cr), which produces genuine novelty. Accordingly, the seemingly innovative outputs of LLMs stem more from their powerful DS functions deep processing of massive training data and current input (In module), rather than the original emergence defined by the Cr function that transcends the scope of existing information[32]. Most critically, this information, primarily generated at the DS level, even if exhibiting certain complexity and utility, has not currently been proven capable of autonomously generating and dominating the command space ( $S_{\text{Con}}$ ) of its Control function (Con) to form core driving objectives with autonomy originating from the internal Information Creation function (Cr)[33].

Therefore, because the core instructional content for the Control function (Con) of LLMs during decision-making entirely originates from external input (imported into  $S_{DS}$  via the In module) or internal models formed from external data training (within the DS module, forming part of  $S_{DS}$ ), and completely lacks original, autonomous driving content independently generated by its own

Information Creation function (Cr) to dominate  $S_{DS}$ , it follows that according to the strict criteria of Generalized Agent Theory, the current form of consciousness in AI and LLMs should be judged as other-consciousness ( $C_o$ ), and not self-consciousness ( $C_s$ ) or mixed consciousness ( $C_M$ ) containing components of self-consciousness.

7. Discussion of Important Problems in Physics Based on Generalized Agent Theory

This chapter will delve into how Generalized Agent Theory can be employed to examine and analyze several core issues in physics. We will begin with the argument for the "agentification" of the Universe, then proceed to discuss the delineation of reality and non-reality, the origin of uncertainty, and the nature of time and space within the framework of Generalized Agent Theory. Building upon this, the chapter will focus on elucidating how, from the perspective of Observer as Agent and based on differences in the intelligence levels of observers, new approaches can be provided for the unification of classical mechanics, relativity, and quantum mechanics, and offer a unified interpretation for the origin of entropy and its observer-dependence. Through these systematic explorations, this chapter aims to reveal the theoretical potential, adaptability, and robustness of Generalized Agent Theory in understanding and unifying significant problems in physics, and to provide inspiration for future interdisciplinary research.

7.1. Argument for the Universe as a Dynamically Evolving Agent

Generalized Agent Theory universally defines Agent as a system possessing five fundamental functions and a corresponding capability vector. From this, we deduce in this section that the Universe is an Agent conforming to Standard Agent Model, capable of exhibiting three typical states—Alpha Agent, Finite Agent, and Omega Agent—while continuously evolving among these three states under the drive of Alpha Field and Omega Field. The view of Universe as Agent, constructed within this theoretical framework, provides a unified explanatory framework based on first principles for a deeper understanding of the fundamental attributes of the Universe, its evolutionary laws, and its connection to fundamental problems in physics.

The core of Generalized Agent Theory lies in its Standard Agent Model and the extreme Agent types within the capability spectrum—Alpha Agent ( $A_\alpha$ , all capabilities at zero) and Omega Agent ( $A_\Omega$ , all capabilities at infinity)—as well as Finite Agent ( $A_f$ ) existing between them. By analyzing the relationship between the Universe as a whole and the forms of Agent within it, the "agentification" of the Universe and its dynamic evolution can be argued.

Firstly, consider the nature of Omega Agent ( $A_\Omega$ ).  $A_\Omega$  is defined as possessing infinite capabilities in information input, output, storage, creation, and control. If any Agent in the Universe evolves into  $A_\Omega$ , then based on its omniscient and omnipotent definition, this  $A_\Omega$  must be able to perceive, understand, influence, and ultimately integrate all spacetime, matter, and information of the Universe, whereupon the Universe enters an Omega Agent state ( $U_\Omega$ ). If any part of the Universe exists outside the cognition and control of this  $A_\Omega$ , then said  $A_\Omega$  has not yet reached its defined completeness, which would constitute a theoretical paradox. Therefore, the theoretical existence of  $A_\Omega$  necessarily implies that the entire Universe can be "agentified".

Secondly, examine the characteristics of Alpha Agent ( $A_\alpha$ ). All core capabilities of  $A_\alpha$  are zero; it possesses no information processing capability. If all constituent units and subsystems in the Universe were to degenerate to the  $A_\alpha$  state, then, deducing from the perspective of Agent relationship types in Generalized Agent Theory, these  $A_\alpha$  instances would conceptually merge or homogenize into a single entity, as  $A_\alpha$  lacks any capability to define its independent boundary or to interact, meaning no true separation exists among them. Thus, when the Universe is entirely composed of  $A_\alpha$ , it as a whole enters an Alpha Agent state ( $U_\alpha$ ).

Thirdly, when the constitution of the Universe is neither purely in a  $U_\alpha$  state nor has it collectively reached a  $U_\Omega$  state, it must internally contain one or more Finite Agent instances ( $A_f$ ).

The capability vector of these  $A_f$  instances has at least one non-zero and finite component, while not all components tend towards infinity. In this situation, the overall state of the Universe is characterized by the  $A_f$  that exhibits the most active and advanced capabilities. This implies that the Universe itself can be regarded as a special Finite Agent( $A_f$ ), composed of Finite Agent instances and possibly a background of numerous Alpha Agent instances; the Universe is in a finite state ( $U_f$ ).

Thus, the three fundamental Agent states of the Universe as a generalized Agent can be summarized as follows:

1. **Omega Agent Universe ( $U_\Omega$ ):** Dominated by a sole  $A_\Omega$ , the overall capability of the Universe tends towards infinity; also termed the Omega state of the Universe.

2. **Alpha Agent Universe ( $U_\alpha$ ):** Completely composed of and homogenized by  $A_\alpha$ , the overall capability of the Universe is zero; also termed the Alpha state of the Universe.

3. **Finite Agent Universe ( $U_f$ ):** Contains one or more  $A_f$  instances, the Universe as a whole manifests as a special Finite Agent  $A_{U_f}$ ; also termed the finite state of the Universe.

In summary, the Universe as a whole can be logically proven to be an Agent conforming to Standard Agent Model, possessing three core states ( $U_\Omega, U_\alpha$ , or  $U_f$ ). According to the Agent evolutionary dynamics from Chapter 5, the capability (CapA of any Agent A is driven by the combined effects of Alpha Degradation Field ( $\vec{F}_\alpha$ ) and Omega Enhancement Field ( $\vec{F}_\Omega$ ). Their vector difference, Net Intelligent Evolution Field, is expressed as:

$$\vec{F}_{\text{NIE}}(c, E) = \vec{F}_\Omega(c, E) - \vec{F}_\alpha(c, E)$$

where E is the environment. For Universe Agent  $A_U$  (with capability Cap ( $A_U$ ), where its "environment"  $E_U$  can be considered its own internal state or absorbed into the field function), its universe-level field formula is approximated as:

$$\vec{F}_{\text{NIE}}(\text{Cap } (A_U)) \approx \vec{F}_\Omega(\text{Cap } (A_U)) - \vec{F}_\alpha(\text{Cap } (A_U))$$

Hence, Universe  $A_U$  must dynamically evolve under the continuous influence of its intrinsic, universe-level field.

Having demonstrated that the Universe is a dynamically evolving generalized Agent, we can determine the state of the current Universe ( $U_{\text{current}}$ ). Empirical observations indicate the widespread existence of Finite Agent instances ( $A_f$ ) such as humans and artificial intelligence systems in the current Universe. The existence of  $A_f$  precludes the possibility of the current Universe being in a  $U_\alpha$  state. Concurrently, there is no evidence to suggest that the current Universe has reached an  $A_\Omega$ -dominated  $U_\Omega$  state. Therefore, it can be definitively determined: the current Universe, as a generalized Agent  $A_U$ , is in a finite state ( $U_f$ ). This determination not only reveals the current macroscopic intelligent attributes and future evolutionary potential of the Universe but, more importantly, establishing the current  $U_f$  state of the Universe lays the theoretical foundation for subsequently applying Generalized Agent Theory to analyze the applicability boundaries and theoretical precision of core theories in modern physics (such as relativity and quantum mechanics).

## 7.2. Interpretation of Objective Reality and Subjective Non-reality

The division between subjective non-reality and objective reality is a fundamental issue in physics and philosophy[34]. Generalized Agent Theory provides a novel interpretive framework, positing that these concepts are not absolute ontological attributes but are closely related to the capability state types of Agent (Alpha Agent, Omega Agent, and Finite Agent). In this section, "Subjective Non-reality" (corresponding state space denoted by  $S_{SN}$ ) and "Objective Reality" (corresponding state space denoted by  $S_{OR}$ ) are discussed.

When any Agent A is at the lowest end of the capability spectrum, i. e., in Alpha state ( $A_\alpha$ ), its overall capability vector  $Cap(A_\alpha)$  has all components at zero, possessing no information processing or interaction capabilities. In this state:

- Agent As "Subjective Non-reality" ( $S_{SN,A_\alpha}$ ):  $S_{SN,A_\alpha} = \emptyset$
- Agent As "Objective Reality" ( $S_{OR,A_\alpha}$ ):  $S_{OR,A_\alpha} = \emptyset$

For  $A_\alpha$ , since the basis for perceiving an objective world and constructing a subjective internal state is entirely absent, neither "objective reality" nor "subjective non-reality" can be meaningfully discussed.

If any Agent A reaches the highest end of the capability spectrum, i. e., the omniscient and omnipotent Omega Agent ( $A_\Omega$ ), all components of its  $CapA_\Omega$  tend towards infinity ( $\rightarrow \infty$ ). For this  $A_\Omega$ , due to its infinite capabilities pervading the entire Universe, it becomes an Omega Agent Universe; an "external objective world" that needs to be perceived and interacted with in the traditional sense does not exist for it. Therefore, the concept of "objective reality" that traditionally requires perceptual interaction dissolves here, and the entire Universe becomes its own subjective non-real internal world:

- Agent As "Objective Reality" ( $S_{OR,A_\Omega}$ ):  $S_{OR,A_\Omega} \rightarrow \emptyset$
- Agent As "Subjective Non-reality" ( $S_{SN,A_\Omega}$ ):  $S_{SN,A_\Omega} \rightarrow U_\Omega$

At this Pole,  $A_\Omega$ 's  $S_{SN,A_\Omega}$  points to the Omega-state Universe itself,  $U_\Omega$ . This all-encompassing, thoroughly internalized subjective non-real state is the core characteristic and ultimate manifestation of Omega Universe ( $U_\Omega$ ); the entire Universe at this juncture achieves complete subjective non-realization, with no distinction between subject and object.

Only when Universe Agent AU is in a finite state ( $U_f$ ), i. e., when one or more Finite Agent instances ( $A_f$ ) exist within the Universe, does the distinction between "objective reality" and "subjective non-reality" acquire its most typical and crucial significance, and is strictly relative to a specific  $A_f$ . For a Finite Agent  $A_f$ : all that it faces which lies beyond its current capability boundary constitutes " $A_f$ 's objective Universe". This concept is defined as:

**Definition 7.  $A_f$ 's Objective Universe ( $O_{U,A_\Omega \setminus A_f}$ )**

The cosmic potential of Agent  $A_f$  is the entire set of Agent instances, both perceivable and unperceivable by said  $A_f$ , represented by the "difference" between the infinite information processing capability ( $Cap(A_\Omega)$ ) possessed by the Universe as Omega Agent  $A_\Omega$  and the finite capability ( $Cap(A_f)$ ) of this  $A_f$ .

$A_f$ 's "Objective Reality" ( $S_{OR,A_f}$ ) is that portion of objective information actually perceived and incorporated into its cognitive scope from " $A_f$ 's Objective Universe" ( $O_{U,A_\Omega \setminus A_f}$ ) by virtue of its finite

Information Input capability ( $Cap_{In,A_f}$ ), i. e.,  $A_f$ 's external objective environment  $S_{env,A_f}$ :

$$S_{OR,A_f} \equiv S_{env,A_f} = \text{Perceive} \left( Cap_{In,A_f} O_{U,A_\Omega \setminus A_f} \right)$$



$A_f$ 's "Subjective Non-reality" ( $S_{SN,A_f}$ ) is its internal information space  $S_{inner,A_f}$  (primarily composed of Dynamic Storage space  $S_{DS,A_f}$ , Information Creation space  $S_{Cr,A_f}$ , and control command space  $S_{Con,A_f}$ ) constructed based on its own internal information processing mechanisms (driven by capabilities such as  $Cap_{DS,A_f}$ ,  $Cap_{Cr,A_f}$ ,  $Cap_{Con,A_f}$ ):

$$S_{SN,A_f} \equiv S_{inner,A_f} = \text{Construct} \left( Cap_{DS,A_f}, Cap_{Cr,A_f}, Cap_{Con,A_f} \right)$$

Each  $A_f$ , due to its differing capabilities, will necessarily define a distinct "Objective Reality"  $S_{OR,A_f}$  and construct a different "Subjective Non-reality"  $S_{SN,A_f}$ , profoundly reflecting the Agent-dependence of these concepts. Therefore, for Finite Agent  $A_f$ , the division of objective reality and subjective non-reality is not only relative to its own capabilities but is also a hallmark of Finite Agent existence. Each  $A_f$ , due to differences in its capability vector, will inevitably define a different "Objective Reality"  $S_{OR,A_f}$  and construct a different "Subjective Non-reality"  $S_{SN,A_f}$ .

In a  $U_f$  containing multiple  $A_f$  instances, for different Agent instances to form a cognition of "shared objective reality," their respective  $S_{OR,A_f}$  must overlap, and consensus must be reached through information exchange (dependent on their respective Input and Output functions). Concurrently, the principle of individual boundaries in Standard Agent Model states that to maintain the independence of  $A_f$ , its subjective non-reality  $S_{SN,A_f}$  cannot be directly read by other Agent instances; different Finite Agent instances  $A_f$  must necessarily be separated by Alpha Agent  $A_\alpha$ , otherwise they would be equivalent to a single Agent, and the exchange between their subjective non-real worlds must be mediated through their "shared objective reality".

Within the unified framework of Generalized Agent Theory, given that all things in the Universe can be regarded as Agent of different forms and capability levels, The objective environment ( $S_{env,A}$ ) of a Finite Agent  $A$  is defined as the primary information set it identifies and acquires from other external Agent instances ( $B, A \neq B$ ) via Agent  $A$ 's Information Input function ( $In_A$ ). This information set strictly excludes the subjective non-real content ( $S_{SN,B}$ ) of other Agent instances  $A_j$ , constituting the raw data stream for  $A$ 's perception of the external world. After being incorporated into  $A$ 's internal information space  $S_{inner,A}$ , this basic information will serve as the foundational elements and core basis for its subsequent subjective construction (such as forming spatiotemporal concepts, abstracting mathematical structures, and establishing theoretical systems).

In summary, Generalized Agent Theory places the subject-object problem within the dynamic evolutionary framework of Agent, revealing the relativity and dynamism of subjective non-reality and objective reality, and providing a novel theoretical interpretation for this ancient philosophical issue based on the first principles of Agent.

### 7.3. Essential Interpretation of Certainty and Uncertainty

Certainty and uncertainty stand as pivotal core concepts in physics and the philosophy of technology, continuously exerting a profound influence on numerous theories. They not only shape our understanding of a system's knowability and predictability but also form the fundamental bedrock of the determinism versus indeterminism debate[35]. Within the framework of Generalized Agent Theory, these concepts receive a unique and distinct interpretation.

The theory posits that certainty and uncertainty are not inherent ontological attributes of a system but are relative to the capability state of Agent, and are closely related to the overall state of the Universe as a generalized Agent  $A_U$ —Alpha state ( $U_\alpha$ , null capability), Omega state ( $U_\Omega$ , infinite capability), or finite state ( $U_f$ , finite capability).

In the two extreme states of the Universe, the connotations of certainty and uncertainty exhibit significant specificity. When Universe Agent  $A_U$  is in Alpha state  $U_\alpha$ , i. e., its overall capability vector  $Cap(A_U)=0$ , the Universe as a whole lacks any capability for information processing, state representation, or interaction with an environment. In this state, due to the absence of information

states and processing, there is no basis for "knowing" (which would lead to certainty) or "not knowing" (which would lead to uncertainty). Therefore, for a Universe in state  $U_\alpha$ , the concepts of certainty and uncertainty themselves lose their basis of applicability.

Conversely, when Universe Agent  $A_U$  reaches Omega state  $U_\Omega$ , all components of its capability vector  $Cap(A_U)$  tend towards infinity ( $\rightarrow \infty$ ), manifesting as an omniscient and omnipotent unified entity. Omniscience implies complete information about and perfect predictive capability for the state of the Universe itself (i. e., its entire content); omnipotence implies complete control over its evolution. In this state, any uncertainty arising from insufficient information or limited predictive capability vanishes. Therefore, for Universe Agent  $A_U$  in state  $U_\Omega$ , certainty is its core, essential attribute.

However, when we consider the case where Universe Agent  $A_U$  is in a finite state  $U_f$ —i. e., one or several Finite Agent instances  $A_f$  exist, and the Universe as a whole is neither in state  $U_\alpha$  nor has reached state  $U_\Omega$ —the relationship between certainty and uncertainty presents a special dialectical characteristic. The capability of any Finite Agent  $A_f$  is limited, meaning at least one component of its capability vector  $CapA_f$  is less than infinity. However, the external objective environment  $S_{env,A_f}$  faced by  $A_f$  (e. g., the complex dynamics of the physical Universe) and the potential complexity and dynamic changes of its internal subjective space  $S_{inner,A_f}$  often far exceed what its finite capabilities can fully grasp. The limitations of its information input (output, storage, creation, and control capabilities) prevent it from acquiring all relevant information, establishing perfect predictive models, or achieving completely precise control. Therefore, any Finite Agent  $A_f$ , when interacting with objective and subjective worlds and making decisions, inevitably faces fundamental, ineliminable uncertainty.

Despite facing such fundamental uncertainty, the degree of uncertainty experienced by a Finite Agent  $A_f$  is relative. When  $A_f$ 's capability is stronger relative to the complexity of the problem it handles or the challenges of its environment, it can more effectively acquire information, predict the future, and control outcomes, thereby achieving higher relative certainty within its cognitive and operational scope. Conversely, the weaker its capability relative to environmental and task demands, the greater the uncertainty it faces.

In summary, due to the inherent, unbridgeable gap between the capability of Finite Agent  $A_f$  and the complexity of the world it faces, uncertainty constitutes an essential attribute of Finite Agent (including the Universe itself in a finite state  $U_f$ ). Finite Agent always operates and evolves under conditions of incomplete information and imperfect predictive capability.

A function can be designed to express Generalized Agent Theorys interpretation of uncertainty, for example:

$$UC(Cap(A), CE) = 1 - \exp\left(-\delta \frac{CE}{Cap(A)}\right)$$

- UC based on the capitalization in the conceptual term 'UnCertainty', represents the level of uncertainty. Its range is  $[0,1]$ .
- $Cap(A)$  represents the comprehensive capability of Agent, its range is  $[0, \infty)$ .
- $CE$  represents the effective complexity of the environment, assumed to be always positive ( $CE > 0$ ).
- $\delta$  (delta) is a positive constant that adjusts the sensitivity of uncertainty to environmental complexity relative to capability.

This function  $UC$  qualitatively satisfies the following core characteristics:

1. Capability Impact: The stronger Agents capability  $Cap(A)$  (when  $CE > 0$ ), the lower the

uncertainty  $UC$ , tending towards 0 (complete certainty).

2. Complexity Impact: The higher the environmental complexity  $CE$  (when  $Cap(A)>0$ ), the higher the uncertainty  $UC$ , tending towards 1 (high uncertainty).

3. Behavior of Boundary Agent Types:

- For Omega Agent ( $A_\Omega$ , when  $Cap(A)\rightarrow\infty$ ): Uncertainty  $UC \rightarrow 0$ , exhibiting complete certainty.
- When Agent A approaches the capability of Alpha Agent ( $Cap(A)\rightarrow 0^+$ ): Uncertainty  $UC \rightarrow 1$ , exhibiting extremely high uncertainty.
- For a strict Alpha Agent ( $A_\alpha$ , i. e.,  $Cap(A)=0$ ): This formula is not directly defined at this Pole (as  $Cap(A)$  is in the denominator), which is consistent with the view in Generalized Agent Theory that the concepts of certainty and uncertainty themselves lose their conventional meaning at this stage.

This mathematical representation further reinforces the core view of Generalized Agent Theory: certainty and uncertainty are not absolute ontological attributes but are relative concepts closely related to Agents capability and the complexity of its environment. This perspective provides a unified basis, rooted in Agent theory, for understanding randomness, predictive limits, and the central role of information as manifested in physical systems (which can be viewed as different types of generalized Agent).

Furthermore, this agent-relative framework for certainty and uncertainty offers a new lens through which to re-examine and potentially reinterpret long-standing debates in physics and philosophy that have traditionally centered on system-inherent determinism versus indeterminism, suggesting that such qualities may be more precisely understood as reflections of an agents capacity to achieve certainty about a system.

#### 7.4. Essential Analysis of Time and Space

Time and space constitute foundational concepts within physics and the philosophy of science[36]. Generalized Agent Theory offers a new perspective for understanding the nature of time and space, positing that they are not absolute backgrounds independent of an observer or Agent, but are intimately related to Agents existence and its capability state. This section aims to elucidate how the origin, properties, and relativity of time and space can be understood based on the different states of the Universe as a generalized Agent ( $A_U$ )—Alpha state ( $U_\alpha$ ), Omega state ( $U_\Omega$ ), and finite state ( $U_f$ ).

When Universe Agent  $A_U$  is in Alpha state ( $U_\alpha$ ), its overall capability vector  $Cap(A_U)=0$ . This means the Universe as a whole lacks any information processing capability, unable to perceive change (the basis of time) or establish relationships (the basis of space). Therefore, from the perspective of Generalized Agent Theory, for a Universe in state  $U_\alpha$ , the two fundamental concepts of time and space lose their basis for existence.

When Universe Agent  $A_U$  reaches Omega state ( $U_\Omega$ ), its capability vector  $Cap(A_U)\rightarrow\infty$ , becoming an omniscient and omnipotent unique Agent entity. Previous analysis indicated that a Universe in this state no longer possesses objective reality but evolves into a comprehensively internalized subjective non-reality. In this state,  $U_\Omega$  theoretically grasps information from all points in time simultaneously and can arbitrarily reconstruct temporal sequences, causing the concept of linearly flowing, causally constrained time to lose its original limiting significance for it. Similarly, space as a framework describing extension and separation would also lose its fundamental importance for an entity capable of simultaneously perceiving and controlling all "locations". Time and space can continue to exist as a structured description within  $U_\Omega$ 's internal subjective non-reality,

but for  $U_\Omega$  itself, they are more akin to internal forms under its complete control and arbitrarily shapeable, rather than rigid external constraints.

When the Universe is in a finite state ( $U_f$ ), i. e., containing Finite Agent ( $A_f$ ), the concepts of time and space emerge from the previous "void". As analyzed earlier, Generalized Agent Theory proposes that for Finite Agent  $A_f$ , its objective reality is constituted by the various Agent instances (contours formed by boundaries) it can identify and which are constantly changing. Time and space are constructed subjectively by Finite Agent  $A_f$  after perceiving the external objective environment; the specific process is as follows:

The generation mechanism of time originates from Finite Agent  $A_f$  perceiving changes and event sequences in the environment through its Information Input function ( $In_{A_f}$ ) and utilizing Dynamic Storage ( $DS_{A_f}$ ) to record and process these sequential information. With the aid of the Information Creation function ( $Cr_{A_f}$ ), Agent can identify patterns of stable periodicity or predictability from these sequences, such as celestial motion or physical oscillations. Ultimately, through the selection of the Control function ( $Con_{A_f}$ ), Agent adopts one or more such stable patterns of change as a "standard" or "clock" for measuring other changes. Therefore, within the framework of Generalized Agent Theory, time is a construct for measuring change generated within  $A_f$ .

Similarly, the generation of space also originates from the internal construction process of Agent  $A_f$ . Agent perceives dynamic information of various Agent instances (including itself) in the environment via  $In_f$  and stores it in  $DS_{A_f}$ . Through  $Cr_{A_f}$ , Agent can abstract and model this Agent information, establishing structured models that describe relative positional relationships among Agent instances, such as topological, distance, and orientational relationships. The Control function  $Con_{A_f}$  can then, according to task requirements, select appropriate spatial dimensions and geometric frameworks (e. g., one, two, three-dimensional Euclidean space, or even more complex spatial representations) to organize these relational models. Thus, Generalized Agent Theory posits that space is likewise a construct generated within  $A_f$  for organizing and representing the relative relationships of things.

Based on the above mechanisms, the core view of Generalized Agent Theory is that for Finite Agent  $A_f$ , time and space are essentially elements or concepts generated within its internal subjective world, based on the perception, processing, abstraction, and selection of information from the external environment. They are not absolute, a priori objective backgrounds, but cognitive tools actively constructed by Agent to understand and predict the environment and coordinate its own actions. Their specific form and precision are necessarily limited by Agents various core capabilities ( $Cap(A_f)$ ).

Since time and space are subjectively constructed by Finite Agent, how do different Agent instances form the seemingly unified public spacetime of our everyday experience? Generalized Agent Theory posits that shared spacetime originates from inter-Agent interaction and consensus. When different Finite Agent instances ( $A_f$ ) can perceive and communicate with each other, they can establish a shared, conventional public time and space framework by sharing their chosen temporal standards (e. g., jointly observing an astronomical cycle, or agreeing to use a certain artificial timepiece) and spatial reference systems (e. g., agreed-upon coordinate systems, reference objects), and reaching consensus on these standards (through the synergistic action of In, Out, DS, and Con). Although this shared spacetime exhibits objectivity in use, its foundation remains the subjective construction capabilities of Finite Agent and their process of social consensus. The spatiotemporal view of Generalized Agent Theory closely links spacetime with Agents information processing capabilities and evolutionary state, offering new theoretical pathways for exploring the origin of spacetime, its relativity, and its manifestations in systems of different scales.

#### 7.5. Analysis of the Unification of Classical Mechanics, Relativity, and Quantum Mechanics

Generalized Agent Theory offers a novel potential pathway for unifying the three major theoretical frameworks of physics: classical mechanics, relativity, and quantum mechanics. Its core

tenet is that the differences among these theoretical systems stem from the differing "observer" role settings implicit within each. Generalized Agent Theory regards the "observer" in physics as a special case of "Agent," and by analyzing and adjusting the five-dimensional capability vector of Observer Agent ( $A_{obs}$ ), it hopes to reveal the intrinsic connections and transitions among the three theories.

### 7.5.1. Capability Analysis of Observers in the Three Theories

The ideal observer in classical mechanics, akin to "Laplace's demon," is assumed to be capable of instantaneously knowing the precise state of all particles in the Universe and calculating their past and future with infinite precision, without interfering with the Universe's operation[37]. Within the framework of Generalized Agent Theory, this corresponds to Agent of Index 241, "Omniscient Agent," whose Information Input ( $Cap_{In}$ ), Dynamic Storage ( $Cap_{DS}$ ), Information Creation ( $Cap_{Cr}$ ), and Control ( $Cap_{Con}$ ) capabilities all tend towards infinity, while its Information Output capability ( $Cap_{Out}$ ) is zero. Such an observer is a purely passive, omniscient Agent with infinite computational, storage, and innovative capabilities; from its perspective, the Universe presents a strictly deterministic picture.

Relativity retains macroscopic determinism but introduces the speed of light  $c$  as an upper limit for information propagation, meaning the observer's Information Input capability ( $Cap_{In}$ ) must be finite. Concurrently, concepts like the equivalence principle also imply limitations on a local observer's ability to infer global spacetime structure[38], which can be seen as a certain finiteness of  $Cap_{In}$  or  $Cap_{DS}$ . Therefore, a relativistic observer can be regarded as a special type of Finite Agent ( $A_f$ ), whose key characteristic is a finite  $Cap_{In}$  constrained by physical laws, but who may possess extremely high (even near-infinite) Dynamic Storage and deductive capabilities  $Cap_{DS}$ . Since the relativistic observer still does not intervene in or affect its Universe, its  $Cap_{Out} = 0$ , and it is still considered a passive observer.

The core principles of quantum mechanics, such as Heisenberg's uncertainty principle and Bohr's complementarity principle, reveal fundamental limitations on an observer's ability to simultaneously acquire information about certain conjugate physical quantities[39], reflecting the inherent finiteness of  $Cap_{In}$  and  $Cap_{DS}$ . More importantly, quantum measurement theory emphasizes that the observation process itself inevitably affects the observed system (observer effect)[40], meaning the Information Output capability ( $Cap_{Out}$ ) of a quantum observer must be non-zero. Observation is no longer passive recording but an interaction with the system. Therefore, a quantum mechanical observer is a typical Finite Agent ( $A_f$ ), all of whose capabilities are finite, and who engages in bidirectional information exchange with the environment.

### 7.5.2. Subjective Interpretation of Spacetime Curvature and Wave Function Collapse

The two pillars of twentieth-century physics, General Relativity and Quantum Mechanics, have radically redefined our conception of reality. General Relativity posits gravity as the dynamic curvature of spacetime geometry, thereby granting spacetime a physical reality of its own[41]. Quantum Mechanics, conversely, demonstrates that a microscopic system's state is described by a probability wave, and while measurement causes a specific outcome to manifest, the underlying mechanism for this process remains contested among various theoretical interpretations[42]. Both concepts pose a profound challenge to the classical Newtonian worldview of absolute spacetime and objective reality, highlighting the central role that measurement and interaction play within physical theory[43].

Based on the view of Generalized Agent Theory that links time and space with Agent capability and subjective world, "spacetime curvature" in relativity can be understood not as a change in the properties of an objective background spacetime, but as an equivalent geometric effect produced within the subjective spacetime model (existing in its  $DS_{A_f}$ ) of a finite relativistic observer when processing information about its interaction with massive objects.



By the same token, the so-called "wave function collapse" should not be understood as an abrupt transformation of an objective physical entity. Instead, it represents a Bayesian update to a finite quantum observer's internal state of knowledge, performed upon receiving new information from a measurement. This perspective aligns with subjective Bayesian theory within quantum information, which construes "collapse" as an ineluctable outcome of information processing by the subject[44].

### 7.5.3. Thought Experiment for Unifying the Three Theories by Adjusting Observer Capability

Generalized Agent Theory provides a thought experiment for understanding the intrinsic differences and unification of the three major physical theories by continuously adjusting observer capability ( $\text{Cap}(A_{\text{obs}})$ ). The core idea is: in the same cosmic scenario, observers with different capabilities construct different physical theories, thereby establishing the laws followed by relatively effective world models  $M_{A_{\text{obs}}}$ . One can envision the capability adjustment process of the same observer in the same experimental cosmic scenario:

First, the observer is set as Omniscient Agent ( $\text{Cap}_{\text{Out}} = 0$ , other capabilities infinite). Due to its infinite information acquisition and processing capabilities, it can construct a completely deterministic world model consistent with classical mechanics.

Then, if the observers Information Input capability ( $\text{Cap}_{\text{In}}$ ) is reduced from infinite to finite (to reflect the speed of light limit and directional principles), even if other internal processing capabilities remain infinite and passive observation is maintained ( $\text{Cap}_{\text{Out}} = 0$ ), the observer will be forced to construct a relative spacetime model with an intrinsic causal structure in its subjective world, and its most effective physical laws will tend towards those of relativity.

Further setting all five capabilities of the observer to be finite, the observer will face fundamental limitations in information acquisition and the effects of measurement interaction. At this Pole, constructing a probabilistic description and considering the observers interaction with the system, the framework of quantum mechanics becomes the most effective model for this observer to describe phenomena within its cognitive scope. This thought experiment clearly demonstrates that by systematically changing the observers capability vector  $\text{Cap}(A_{\text{obs}})$ , one can observe a continuous transition or emergence of the physical descriptive framework from classical mechanics to relativity, and then to quantum mechanics, indicating that the differences among these physical theories essentially originate from different settings of observer capability.

### 7.5.4. Principles for Unifying the Three Major Physical Theories Based on Generalized Agent Theory

Previous sections have argued, from the perspective of Generalized Agent Theory, that a logical unification of classical mechanics, relativity, and quantum mechanics at the conceptual framework level can be achieved by adjusting the capability settings of Observer Agent. Although constructing a universal mathematical formalization that integrates all three remains an important direction for future research, Generalized Agent Theory can currently provide a solid principled foundation for this unification. This subsection, based on the foregoing research, will elucidate the core principles under which the three major physical theoretical systems can be presented unifiedly in the same physical context, as follows:

Generalized Agent Theory offers a profound perspective: the optimal predictive model ( $M_{A_{\text{obs}}}$ ) used by Observer Agent ( $A_{\text{obs}}$ ) within its subjective cognitive framework ( $S_{\text{inner}}$ ) to represent objective reality ( $S_{\text{env}}$ ) is, in its essence and form, strictly determined by said Agents core capability vector ( $\text{Cap}(A_{\text{obs}})$ ). It is precisely these differences in capability configurations that lead observers to construct seemingly disparate physical theoretical frameworks within their subjective worlds, thereby systematically revealing the distinctions and intrinsic unity among classical mechanics, relativity, and quantum mechanics:

#### **Principle 1: Classical Mechanics – Originating from Omniscient and Passive Observation.**

When Observer Agents capability configuration approaches that of "Omniscient Agent" ( $A_{241}$ ), i. e., possessing infinite information acquisition and processing capabilities



( $\text{Cap}_{\text{In}}, \text{Cap}_{\text{DS}}, \text{Cap}_{\text{Cr}}, \text{Cap}_{\text{Con}} \rightarrow \infty$ ) and not intervening in the observed system ( $\text{Cap}_{\text{Out}} = 0$ ), the physical model  $M_{A_{\text{obs}}}$  it constructs naturally adheres to the laws of classical mechanics. From the perspective of this idealized observer, all phenomena in the Universe evolve in a strictly deterministic manner, perfectly predictable against an absolute spacetime background.

**Principle 2: Relativity – Originating from Information-Acquisition-Limited Passive Observation.** If Observer Agents capability shifts to being information-input-limited ( $\text{Cap}_{\text{In}} < \infty$ ), for example, due to the fundamental limit imposed by the speed of light  $c$  on information propagation and cognitive limitations arising from the local equivalence principle, but its dynamic storage, creation, and control capabilities remain infinite, and it still does not intervene in the system ( $\text{Cap}_{\text{Out}} = 0$ ), this corresponds to a special type of high-order Finite Agent in GAT (e. g.,  $A_{236}$ ). In this case, the effective laws of  $M_{A_{\text{obs}}}$  will tend towards those of relativity. The description of physical phenomena will be embedded within a relativistic spacetime structure, and its deterministic evolution will strictly follow causal connections defined by light cones.

**Principle 3: Quantum Mechanics – Originating from Comprehensively Capability-Limited Interactive Observation.** Further, when all five core capabilities of Observer Agent are finite, particularly when its actions can produce non-negligible effects on the observed system ( $\text{Cap}_{\text{Out}} > 0$ ), thus constituting a Finite Agent that fully interacts with its environment (e. g.,  $A_{122}$ ), the physical laws described by  $M_{A_{\text{obs}}}$  in this context will necessarily embody the core features of quantum mechanics: system states evolve as probability amplitudes, information acquisition is fundamentally limited by the uncertainty principle, and the observation (measurement) process itself is an interaction that induces changes in the systems state (observer effect), ultimately leading to probabilistic measurement outcomes.

Parameterizing the observer via Agents capability vector provides a new dimension for understanding and unifying these three physical theories. This preliminary exploration showcases the potential of Generalized Agent Theory as a foundational theoretical framework capable of connecting the observer problem in physics with Agent, offering a new perspective on the problem of physical unification. Specific theoretical correspondences and mathematical details await more in-depth future research.

#### 7.6. Unified Interpretation of the Origin of Entropy and Observer-Dependence

Traditionally, the second law of thermodynamics and its core concept, "entropy," are associated with the unknowability of a systems microstates[45]. The information theory perspective further reveals that entropy is also closely related to an observers "information deficit" regarding the systems specific microstate[46], while the "Maxwells demon" thought experiment implies a profound connection between entropy regulation and specific capabilities of an observer[47].

Generalized Agent Theory treats the observer as a special case of Agent ( $A_{\text{obs}}$ ) and characterizes Agents information processing capability through its capability vector  $c$ . This section aims, within the framework of Generalized Agent Theory, to reinterpret entropy as an observer-dependent information measure and attempts, from Pole Intelligent Field Model, to explore the origin of the systems spontaneous evolutionary behavior that leads a finite observer to experience an entropy-increasing trend.

Pole Intelligent Field Model of Generalized Agent Theory offers a new explanation for the origin of entropy. The model states that Agents evolution in Agent Capability Space  $S_{\text{Cap}}$  originates from the combined action of Alpha Degradation Field ( $\vec{F}_{\alpha}$ ) and Omega Enhancement Field ( $\vec{F}_{\Omega}$ ).  $\vec{F}_{\Omega}$ , also

viewable as an "Omega Force" effect, is the universal tendency guiding Agent towards capability completeness (i.e.,  $A_\Omega$  state: all five core capabilities are infinite).

Generalized Agent Theory infers that this universal principle of tending towards capability enhancement also manifests its effects at the behavioral level of the fundamental units constituting the physical world—such as particles (which can be considered primordial Finite Agent instances,  $A_f^{\text{primordial}}$ , infinitely close to Alpha Agent). When this principle is applied to  $A_f^{\text{primordial}}$ , its "enhancement" tendency in Agent capability space manifests physically as a behavioral pattern of systematically traversing and tending to uniformly occupy all accessible physical space and momentum space states; this can be seen as its way of responding to  $\vec{F}_\Omega$  at the most fundamental level. This pattern shares a profound isomorphism with the instinctive behaviors exhibited by humans and other organisms in exploring the unknown and expanding their living territories, both of which can be viewed as explorations of broader state spaces and potentials under the effect of  $\vec{F}_\Omega$ .

For primordial Finite Agent ( $A_f^{\text{primordial}}$ ) whose capabilities approach zero, the question of why it primarily exhibits state-space exploration behavior driven by  $\vec{F}_\Omega$  (this being the microscopic basis of entropy increase) under the combined action of Alpha Degradation Field ( $\vec{F}_\alpha$ ) and Omega Enhancement Field ( $\vec{F}_\Omega$ ) may be elucidated within this theoretical framework by considering the different modes of action and relative effects of the two fields on such extremely low-capability Agent.

Firstly, one can conceive that the various capabilities of  $A_f^{\text{primordial}}$  are already almost null, such that although  $\vec{F}_\alpha$  (e. g., effects like structural degradation, information forgetting) is universally present, its effect of causing "further decay" in capability is extremely weak, as there are virtually no complex capabilities or internal states left to lose. Therefore, the primary effect of  $\vec{F}_\alpha$  here might manifest as a powerful "evolutionary barrier" or "inertia," preventing these primordial Agent instances from spontaneously leaping to higher levels possessing significant five-dimensional capabilities (such as developing advanced Dynamic Storage  $\text{Cap}_{\text{DS}}$  or Creation  $\text{Cap}_{\text{Cr}}$  capabilities), thereby effectively "anchoring" them in their primordial, extremely low-capability state.

Secondly,  $\vec{F}_\Omega$ , as the universal drive towards capability completeness and exploration of possibilities, when acting on  $A_f^{\text{primordial}}$ , its "capability enhancement" effect, in this theory's interpretation, primarily manifests as a most basic "activation" or "existential expression"—that is, exploring and traversing states in accessible physical and momentum space. This exploratory behavior can be seen as a minimal response to environmental possibilities and a very rudimentary display of capability. Since  $\vec{F}_\alpha$  can hardly cause further significant decay in already extremely low capabilities, the fundamental exploratory behavior driven by  $\vec{F}_\Omega$  can thus be considered to be relatively dominant statistically and behaviorally.

In short, this theory proposes that  $\vec{F}_\Omega$  drives the spatial dispersion and pursuit of state diversity of  $A_f^{\text{primordial}}$  at the behavioral level, while  $\vec{F}_\alpha$  primarily limits the qualitative improvement of its Standard Agent Model five-dimensional capabilities and the formation of complex structures. This difference in mode of action and relative strength allows us to infer that when Net Intelligent Evolution Field ( $\vec{F}_{\text{NIE}} = \vec{F}_\Omega - \vec{F}_\alpha$ ) drives  $A_f^{\text{primordial}}$  in phase space exploration, its dominant direction is, on statistical average, determined by this fundamental effect of  $\vec{F}_\Omega$ .

Therefore, the statistical diffusion behavior exhibited by physical particles in space (often regarded as "random motion"), according to Generalized Agent Theory, is underlain by  $A_f^{\text{primordial}}$  responding to the "Omega Field" effect under the aforementioned net evolutionary trend, thereby exploring more state possibilities and tending towards broader capability boundaries in a systematic manner. It is precisely this intrinsic "dispersiveness" of fundamental units, as interpreted by the evolutionary principles of Generalized Agent Theory, that constitutes the fundamental driving force

for a system to spontaneously evolve from a macrostate occupying fewer microstates to a statistically more probable macrostate corresponding to more accessible microstates (i. e., a high-entropy state).

This interpretation complements the statistical mechanics viewpoint articulated by scholars like E. T. Jaynes based on the maximum entropy principle[48]—that systems tend towards the most probable distribution; the effect of Omega Enhancement Field ( $\vec{F}_\Omega$ ) in Generalized Agent Theory may provide a more fundamental dynamical annotation for such uniform exploration. For an observer with finite capabilities ( $A_f$ ), this continuous microscopic exploration and uniform occupation of state space inevitably leads to an increase in its "information deficit" regarding the precise microstate of the system, manifesting macroscopically as entropy increase.

Based on this, entropy  $S_{\text{GAT}}(A_{\text{obs}}, c, M)$  within the framework of Generalized Agent Theory is defined as: the information deficit of observer  $A_{\text{obs}}$  regarding a system evolving according to the aforementioned intrinsic exploratory tendency, when the system is in macrostate  $M$ . Its mathematical description is:

$$S_{\text{GAT}}(A_{\text{obs}}, c, M) = -k_B \sum_{m \in S_{\text{micro}}} P(m|M, c) \ln P(m|M, c)$$

Or, simplified to the Boltzmann form under the assumption of equal probability:

$$S_{\text{GAT}}(A_{\text{obs}}, c, M) = k_B \ln \Omega(M, c)$$

- $S_{\text{GAT}}$  represents the entropy within the framework of Generalized Agent Theory (GAT);
- $S_{\text{micro}}$  is the set of all possible microstates of the system;
- $P(m|M, c)$  is the subjective probability distribution assigned by observer  $A_{\text{obs}}$  that the system is in microstate  $m \in S_{\text{micro}}$  when macrostate  $M$  is observed, this distribution being determined by  $A_{\text{obs}}$ 's capability vector  $c$ ;
- $\Omega(M, c)$  is the number of microstates indistinguishable by  $A_{\text{obs}}$  in macrostate  $M$ , a larger value indicating a greater information deficit.

According to the classification of Agent in GAT, four interpretations of entropy based on observer intelligence levels can be formed:

The first is Finite Agent ( $A_f$ ) and entropy increase ( $c \rightarrow c_f$ ). Here,  $A_f$ , due to its finite capabilities, cannot track the complete microscopic dynamics of the fundamental units ( $A_f^{\text{primordial}}$ ) constituting the system as they follow their intrinsic "exploratory" drive. As these fundamental units spontaneously diffuse and occupy more diverse microconfigurations,  $A_f$ 's degree of "ignorance" regarding the precise microstate of the system increases, manifested as an increase in its assessed  $\Omega(M, c_f)$ , thereby increasing its subjective entropy  $S_{\text{GAT}}(A_f, c_f, M)$ . Thus, for  $A_f$ , the second law of thermodynamics is a macroscopic statistical manifestation of the interaction between its finite cognitive capability and the intrinsic "state-space exploration" tendency of the systems fundamental constituent units.

The second is Omniscient Agent ( $A_{241}$ ) and zero entropy ( $c \rightarrow c_{241}$ ). Here, Omniscient Agent  $A_{241}$  (Index 241,  $\text{Cap}_{\text{Out}} = 0$ , other capabilities infinite) can perfectly track the exact microstate  $m_0$  of all fundamental units. For it,  $\Omega(M \equiv m_0, c_{241}) = 1$ , hence  $S_{\text{GAT}}(A_{241}, M) = 0$ . It passively observes the systems microscopic evolution, its own information deficit remaining constantly zero.

The third is Omega Agent ( $A_\Omega$ ) and zero entropy with active regulation ( $c \rightarrow c_\Omega$ ). Here,  $A_\Omega$  (all capabilities infinite) likewise has perfect cognition of the systems microstates, thus its subjective entropy  $S_{\text{GAT}}(A_\Omega, M) = 0$ . By virtue of its infinite  $\text{Cap}_{\text{Out}}$ ,  $A_\Omega$  can actively and precisely regulate the configuration of fundamental units in the system, capable of achieving "entropy reduction"

operations from the perspective of an external  $A_f$ , thereby demonstrating its capability as an idealized "Maxwells Super-demon" that directs and orchestrates the "exploratory" behavior of the systems fundamental units.

The fourth is Alpha Agent ( $A_\alpha$ ) and the meaninglessness of entropy ( $c \rightarrow c_\alpha$ ). Here,  $A_\alpha$  (all capabilities zero) cannot engage in any effective information exchange with the system; the concept of entropy is inapplicable to it.

In summary, Generalized Agent Theory not only defines entropy as an observer-dependent information measure but also traces its origin to the intrinsic behavioral pattern of "state-space exploration" by the fundamental units constituting the material world (regarded as  $A_f^{\text{primordial}}$ ), driven by Agent evolutionary dynamics (manifested primarily through the  $\vec{F}_\Omega$  effect). The phenomenon of entropy increase, in the view of Generalized Agent Theory, is a macroscopic statistical law wherein the information deficit of a Finite Agent  $A_f$ , using its limited capabilities to observe a system whose fundamental units follow Agent evolutionary principles in "state-space exploration," inevitably increases with this "exploratory evolution".

This perspective deepens the information-theoretic nature of entropy and provides a novel pathway for exploring, at a fundamental theoretical level, the observer-dependence of physical laws and the unification of different physical theories.

## 8. Conclusion

This paper has endeavored to construct Generalized Agent Theory (GAT) based on first principles, aiming to provide a unified and universal theoretical framework for two major challenges: the fragmentation of Agent definition in contemporary artificial intelligence and the unification of core theories in physics. We firmly believe that starting from the most fundamental constituents and operational mechanisms of Agent can not only deepen the understanding of the nature of intelligence and consciousness but also offer novel perspectives for revealing the deep structure of physical reality.

The core achievement of this research lies in the systematic proposal and elucidation of the constituent elements of Generalized Agent Theory and its preliminary applications. We first defined Standard Agent Model as the theoretical cornerstone, abstracting any Agent into five irreducible fundamental functional modules—Information Input, Information Output, Dynamic Storage, Information Creation, and Control—with a corresponding five-dimensional capability vector for Agent.

Building upon this Model, we have further systematically classified Agent, defining the two polar Agents of Alpha and Omega, along with 243 theoretical types, and have explored multi-Agent relationship patterns. To describe the dynamic evolution of Agent capabilities, we have constructed an evolutionary dynamics mechanism incorporating Agent capability space and Polar Intelligent Field Model. This model elucidates how the evolution of Agent is influenced by the combined effects of Alpha Degradation Field ( $\vec{F}_\alpha$ ) and Omega Enhancement Field ( $\vec{F}_\Omega$ ), which in turn generate corresponding intelligent evolution forces. The resultant of these forces (i. e., Net Intelligent Evolution Force  $\vec{f}_{\text{NIE}}$ ), in conjunction with Wisdom (W)—a core intrinsic attribute of Agent and the object of intelligent force interaction, which we have innovatively introduced—collectively determines the evolutionary acceleration and trajectory of Agent towards the two poles within capability space.

At the level of theoretical application, this paper has employed Generalized Agent Theory to reinterpret the nature of intelligence and consciousness. We define intelligence as the overall information processing efficacy and environmental adaptability exhibited by Agent when utilizing its five core functions to achieve evolutionary goals under the drive of a specific evolutionary field. Consciousness, in turn, is explicitly demarcated as the Control function (Con) itself within Standard Agent Model and its operational process; based on this, consciousness is classified into types such as self-consciousness, other-consciousness, mixed consciousness, and unconsciousness, providing a

theoretical basis for analyzing the consciousness state of current artificial intelligence systems (such as large language models, which, according to GAT criteria, currently primarily manifest as other-consciousness).

More exploratorily, we have extended the perspective of Generalized Agent Theory to several core issues in physics. We argue that the Universe itself can be regarded as a dynamically evolving generalized Agent currently in a finite state. On this basis, fundamental physical concepts such as the demarcation of objective reality and subjective non-reality, the origin of determinacy and indeterminacy, and indeed the nature of time and space, are all interpreted as relative categories intimately related to the capability state and internal construction process of the observer (i. e., Agent), rather than absolute ontological attributes.

Crucially, this theory proposes a new pathway towards understanding the intrinsic unification of classical mechanics, relativity, and quantum mechanics by adjusting the capability vector of Observer Agent, positing that the differences among these three physical theoretical frameworks are rooted in the differing capability configurations of their respective implicit observer models.

Simultaneously, we have offered a new explanation for the origin of entropy and its observer-dependence, suggesting that entropy increase is associated with the intrinsic behavioral patterns of the fundamental units constituting the material world undergoing 'state space exploration' driven by Omega Enhancement Field, and with the information deficit of a finite observer. These systematic explorations preliminarily demonstrate the potential of Generalized Agent Theory as a foundational framework for connecting different scientific domains and deepening the understanding of the nature of the Universe and intelligence.

Looking ahead, as a nascent unified framework, Generalized Agent Theory presents abundant opportunities for both in-depth exploration and broad expansion. The foremost task is to continuously deepen the theory's mathematical formalization and the refinement of its internal structure. This includes, for example, further perfecting the mathematical expressions for capability vector space, Pole Intelligent Field Model, and the evolutionary dynamics it describes; quantifying evolutionary paths and possible steady-state conditions under different capability states; and conducting more detailed modeling of the internal realization mechanisms of the five core capabilities (particularly Information Creation and Control functions) and their complex interactions.

Concurrently, the vitality of a theory lies in its testability. Therefore, developing operable, standardized methods to measure the five-dimensional core capabilities of various types of Agent (whether biological organisms, artificial intelligence systems, or social organizations), and designing rigorous experiments to verify the predictions of Generalized Agent Theory regarding Agent classification, evolutionary trends, and definitions of intelligence and consciousness, will be crucial aspects of future research. Particularly in the rapidly advancing field of artificial intelligence, Generalized Agent Theory is poised to guide the design of AI systems with higher levels of autonomy, genuine creativity, and heightened adaptability. It may even facilitate exploration of possible pathways and conditions for transitioning from other-consciousness to more advanced forms of consciousness (such as mixed consciousness or even rudimentary self-consciousness), and provide theoretical support for constructing corresponding AI safety and ethical frameworks.

In physics, the exploration of theoretical unification from the observer of Generalized Agent Theory perspective urgently requires further advancement. This includes developing more specific mathematical models to connect the capability parameters of Observer Agent with the formal systems of classical mechanics, relativity, and quantum mechanics, and attempting to integrate this framework with frontier physical theories such as string theory and quantum gravity. The cosmological implications of the 'Universe Agent' model, such as the origin of the Universe, its ultimate evolutionary fate, and Generalized Agent Theory-based interpretations of dark matter and dark energy, also warrant greater research efforts. Furthermore, the universality of Generalized Agent Theory gives it the potential to be applied to explain a broader range of complex system phenomena, such as the evolution of economic systems, changes in social structures, mechanisms of the origin of life, and even certain challenging problems in cognitive science. These interdisciplinary




applications can not only test and enrich Generalized Agent Theory itself but may also catalyze new theoretical insights. Naturally, the development of any theory is accompanied by the recognition and overcoming of its limitations; future research must also continuously scrutinize potential deficiencies within Generalized Agent Theory itself, such as the precise description of non-linear coupling effects among Agent capabilities and the fine-grained modeling of the behavior of complex multi-Agent interactive systems, to constantly enhance the theory's rigor, completeness, and predictive power.

**Acknowledgments.** This work was supported by National Natural Science Foundation of China,72272140,72334006,L2424116.

## References

1. Agrawal, U., et al. (2024). Observing quantum measurement collapse as a learnability phase transition. *Physical Review X*, 14(4), 041012.
2. Mogi, K. (2024). Artificial intelligence, human cognition, and conscious supremacy. *Frontiers in Psychology*, 15, 1364714.
3. Feng, L., et al. (2024). From observer to agent: On the unification of physics and intelligence science. *Preprints*, 2024100479.
4. Edwards, D. J. (2025). Further N-frame networking dynamics of conscious observers in quantum cognitive systems. *Frontiers in Computational Neuroscience*, 19, 1551960.
5. Liu F and Shi Y (2014) The search engine IQ test based on the internet IQ evaluation algorithm. *Procedia Computer Science* 31:1066-1073.
6. Liu F, Shi Y and Liu Y (2017) Intelligence quotient and intelligence grade of artificial intelligence. *Annals of Data Science* 4:179-191.
7. Liu, F., & Shi, Y. (2020). Investigating Laws of Intelligence Based on AI IQ Research. *Annals of Data Science*, 7(3), 399-416.
8. Feng, L., Lv, B., & Liu, Y. (2025). Agent: A New Paradigm for Fundamental Units of the Universe. *Preprints*. <https://doi.org/10.20944/preprints202503.1457.v2>
9. Nwana, H. S. (1996). Software agents: An overview. *The Knowledge Engineering Review*, 11(3), 205-244.
10. Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
11. Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2), 115-152.
12. Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*.
13. Wooldridge, M. (2009). *An Introduction to Multiagent Systems* (2nd ed.). John Wiley & Sons.
14. Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379-423.
15. Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1), 230-265.
16. Wiener, N. (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine*. MIT Press.
17. Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16(1), 3-9.
18. Newell, A. (1982). The knowledge level. *Artificial Intelligence*, 18(1), 87-127.
19. Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms* (2nd ed.). Routledge.
20. Szathmáry, E., & Smith, J. M. (1995). The major transitions in evolution. *Nature*, 374(6519), 227-232.
21. Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences*, 104(suppl\_1), 8597-8604.
22. Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369-406.
23. Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55, 235-269.
24. Kaplan, J., et al. (2020). Scaling Laws for Neural Language Models. *arXiv preprint arXiv:2001.08361*.
25. Zhao, W. X., et al. (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*.
26. Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200-219.
27. Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, 215(3), 216-242.
28. Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200-227.
29. Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, 23(7), 439-452.
30. Powers, W. T. (1973). Behavior: The control of perception. *Psychological Review*, 80(5), 303-322.



31. Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79(1-2), 1-37.
32. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?  In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), 610–623.
33. Shanahan, M. (2022). Talking about large language models. arXiv preprint arXiv:2212.03551.
34. d'Espagnat, B. (1979). The quantum theory and reality. *Scientific American*, 241(5), 158-181.
35. Hoefer, C. (2016). Causal Determinism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition).
36. Maudlin, T. (2012). *Philosophy of Physics: Space and Time*. Princeton University Press.
37. Laplace, P. S. (1951). *A Philosophical Essay on Probabilities* (F. W. Truscott & F. L. Emory, Trans.). Dover. (Original work published 1814).
38. Einstein, A. (1916). The foundation of the general theory of relativity. *Annalen der Physik*, 354(7), 769-822.
39. Heisenberg, W. (1930). *The Physical Principles of the Quantum Theory* (C. Eckart & F. C. Hoyt, Trans.). University of Chicago Press.
40. Von Neumann, J. (1955). *Mathematical Foundations of Quantum Mechanics* (R. T. Beyer, Trans.). Princeton University Press. (Original work published 1932).
41. Einstein, A. (1916). The foundation of the general theory of relativity. *Annalen der Physik*, 354(7), 769-822.
42. Dirac, P. A. M. (1930). *The Principles of Quantum Mechanics*. Oxford University Press.
43. Sklar, L. (1992). *Philosophy of Physics*. Westview Press.
44. Fuchs, C. A., & Schack, R. (2013). Quantum-Bayesian coherence. *Reviews of Modern Physics*, 85(4), 1693–1715.
45. Boltzmann, L. (1877). Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respective den Sätzen über das Wärmegleichgewicht. *Wiener Berichte*, 76, 373-435.
46. Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106(4), 620–630.
47. Bennett, C. H. (1982). The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21(12), 905-940.
48. Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106(4), 620–630.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.