Rethinking Intelligence in the Era of Large Language Models

Guangming Shi, Ning Lan, Xuemei Xie, Jinjian Wu

***Abstract*—In the information age, individuals and organizations have accumulated massive volumes of digital data. As society enters the intelligence era, the primary demand has shifted toward effectively utilizing these data through intelligent models. However, prevailing large language models (LLMs) typically require data to be uploaded to centralized cloud servers for processing, making it difficult to ensure data privacy. Consequently, users increasingly seek small, locally deployable models that can operate directly on private data. To realize this expectation, this paper explores intelligent systems built upon small models rather than monolithic large models. On this basis, we first revisit the essence of intelligence as the capability to utilize knowledge to solve problems and accomplish tasks, and further propose the first principle of intelligence, which models intelligence as a closed-loop, goal-oriented process. Within this process, an agent, under external constraints, leverages knowledge to evaluate discrepancies between its internal state and task objectives, formulates strategies to reduce these discrepancies, and iteratively executes actions to converge toward goal completion. From this perspective, intelligence is viewed as a collaborative system of interdependent cognitive functions. Grounded in this theoretical principle, we propose the six-capability network, a knowledge-driven cognitive architecture that decomposes intelligence into six fundamental capabilities: observation, attention, understanding, discrimination, memory, and execution. These capabilities constitute the core cognitive model of the proposed framework and are each realized by lightweight, deployable small models operating over structured knowledge representations. Finally, to demonstrate the executability of the network, we consider a networked intelligence system. In this system, agents exchange semantic symbols that encode knowledge rather than raw data, enabling each agent to operate within its functional scope while acquiring missing knowledge from others when needed. This approach offers a new intelligence paradigm, enabling deployment in private, local, and resource-constrained environments.**

***Keywords*—the first principles of intelligence, collaborative intelligence, six-capability network, semantic communication network**

Manuscript received Jan. 21, 2026; revised Mar. 04, 2026; accepted Mar. 09, 2026. This work was supported by the National Key Research and Development Program of China under Grant 2023YFA1008500. The associate editor coordinating the review of this paper and approving it for publication was J. Zhang.

G. M. Shi, N. Lan, X. M. Xie, J. J. Wu. School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: gmshi@xidian.edu.cn; lann@stu.xidian.edu.cn; xmxie@mail.xidian.edu.cn; jinjian.wu@mail.xidian.edu.cn).

G. M. Shi. Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: gmshi@xidian.edu.cn).

X. M. Xie. Guangzhou Institute of Technology, Xidian University, Guangzhou 510555, China (e-mail: xmxie@mail.xidian.edu.cn).

X. M. Xie, J. J. Wu. Pazhou LAB-Huangpu, Guangzhou 510700, China (e-mail: xmxie@mail.xidian.edu.cn; jinjian.wu@mail.xidian.edu.cn).

I. Introduction

Human society is undergoing a profound transition from the information age to the intelligence age[1-3]. The defining contribution of the information age lay in the Internet, which connected people, information, and services through standardized protocols and network infrastructures, dramatically reducing the cost of communication and collaboration[1,3]. As society enters the intelligence age, the central objective is no longer merely to connect the world, but to enable the world to be understood. Artificial intelligence (AI) is thus expected to extract meaning from data and support reasoning, marking a shift from information-driven to intelligence-driven social development.

This transformation is accompanied by a fundamental change in user demand. Individuals and organizations now accumulate massive volumes of digital data through long-term operation, research, and daily activities, and increasingly rely on intelligent systems to assist in analyzing and utilizing these data. At the same time, strong requirements have emerged for local deployment, privacy preservation, and controllability, particularly in homes, schools, research institutions, and small or medium-sized enterprises. Users expect intelligent systems to provide reliable reasoning and problem-solving capabilities while minimizing the need to upload raw data to centralized servers. This raises a fundamental question—can

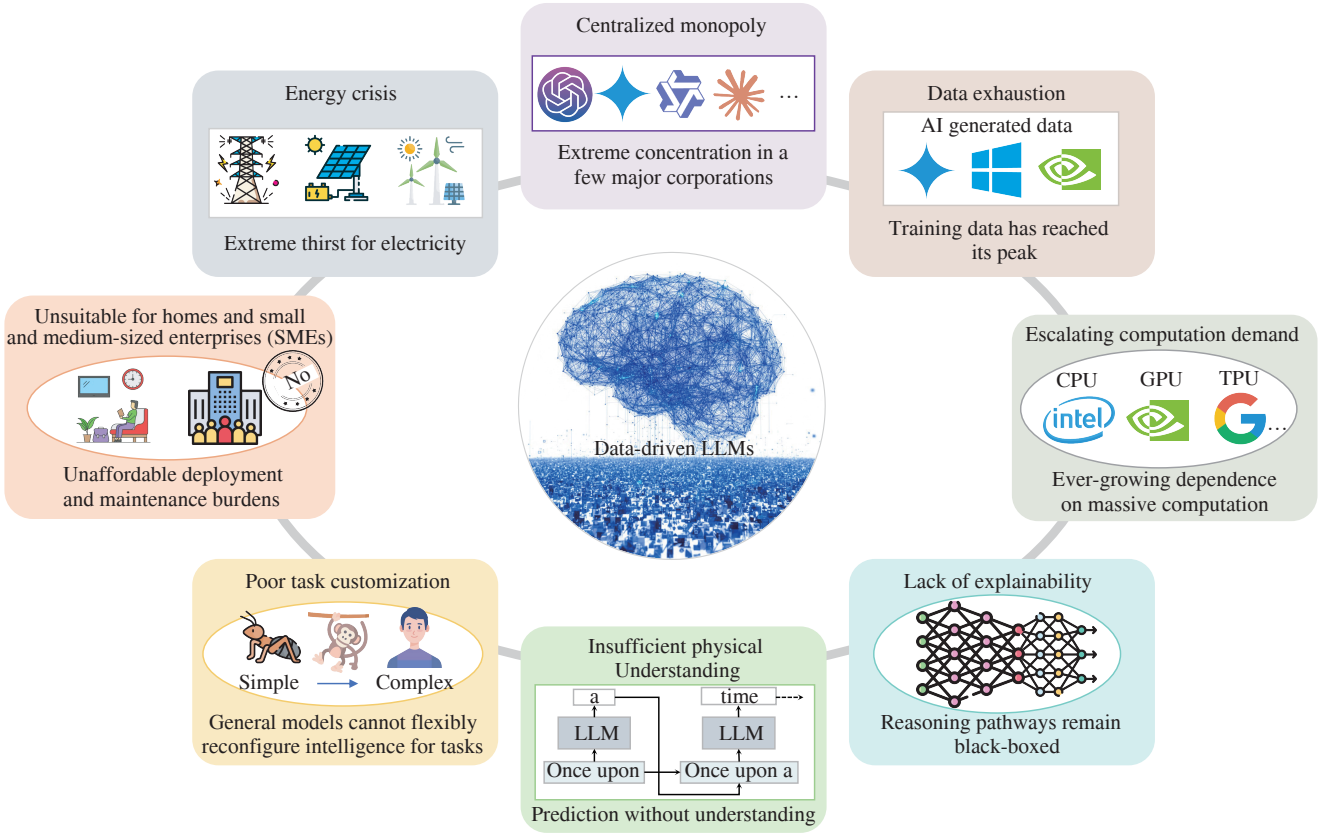

Fig. 1 The current problems and challenges in the era of large language models (LLMs)

the prevailing large-model paradigm truly satisfy these emerging requirements?

In recent years, the intelligence age has been dominated by large-scale models, which have demonstrated remarkable success in language understanding, content generation, and cross-task generalization. From a technical perspective, most contemporary AI systems remain grounded in three classical modeling paradigms: regression, classification, and generative modeling[4-5]. Regression models answer "how much", classification models answer "what it is", and generative models address "how to generate". Despite their differences, these paradigms share a common foundation—learning statistical correlations from large-scale data. While this data-driven approach has yielded impressive empirical performance, it also fundamentally shapes the form of intelligence that such models can achieve.

As emphasized by Academician Zhang Bo at the 2025 Frontier Forum on Large Model Technologies, correlation-based learning, although effective in pattern extraction, faces intrinsic limitations in controllability, reflective cognition, spatiotemporal understanding, and verifiable reasoning. These limitations suggest that the challenges confronting large models are not merely matters of engineering optimization, but reflect deeper constraints imposed by the underlying paradigm. Motivated by this perspective, we systematically analyze the challenges of the large-model paradigm from multiple dimensions, as shown in Fig. 1.

Firstly, from the perspective of application resources, large models impose increasingly severe constraints on energy and infrastructure. Training and inference depend on massive compute clusters, leading to rapidly escalating power consumption and making AI computation a major contributor to global data-center energy demand[4,6-9]. Moreover, only a small number of technology corporations possess the resources required to develop and maintain such systems, resulting in growing centralization of AI capabilities and limiting participation by smaller organizations and academic institutions[2-3,10]. Secondly, from the perspective of application modes, large models are approaching data exhaustion[11]. As high-quality natural data becomes increasingly scarce, reliance on synthetic data has grown substantially. While this alleviates data shortages, it introduces feedback-loop risks, where models are trained on their own outputs, amplifying bias and semantic drift. Thirdly, from the perspective of model capabilities, several fundamental limitations emerge. Expanding parameter scales lead to escalating computational demand[12-13], while the lack of interpretability renders large models opaque black boxes[14-16]. Furthermore, token-based

prediction mechanisms result in insufficient understanding of the physical world[17], and monolithic architectures prevent flexible task customization, causing excessive computational overhead even for simple tasks[12]. Together, these factors render large models difficult to deploy in privacy-sensitive and resource-constrained environments such as homes, schools, and small enterprises[14,18-20].

Collectively, these challenges expose a structural tension in the intelligence age. On one hand, large models offer powerful general-purpose processing capabilities, but rely on centralized, cloud-based computation that requires continuous transmission of raw user data. On the other hand, small models are naturally suited for local deployment and privacy preservation, but their functionality is typically limited, making it difficult for any single model to handle complex reasoning tasks. This contradiction indicates that the difficulty does not stem solely from model scale, but from a deeper misalignment between existing notions of intelligence and the evolving demands of intelligent applications.

To address this misalignment, we revisit the essence of intelligence[17,21-22]. Rather than equating intelligence with scale or data accumulation, we define intelligence as the capability to utilize knowledge to solve problems and accomplish tasks. Based on this definition, we formulate the first principle of intelligence, which characterizes intelligence as a goal-oriented closed-loop process—an agent continuously evaluates discrepancies between its internal state and task objectives, and iteratively selects actions to reduce these discrepancies using available knowledge. Under this formulation, intelligence depends not on how much data is processed, but on how efficiently knowledge is represented, transferred, and utilized.

Guided by this principle, we propose a knowledge-driven collaborative architecture termed the six-capability network, in which intelligence is decomposed into six functional modules—observation, attention, understanding, discrimination, memory, and execution. Each capability is implemented using lightweight, deployable small models operating over structured knowledge representations rather than raw data. When local knowledge proves insufficient, modules selectively query cloud-based large models for high-level guidance, ensuring that what circulates within the system is abstracted knowledge rather than sensitive data. Building on this design, we further construct a networked intelligence system in which agents collaborate exclusively through knowledge-level exchanges, providing a scalable and privacy-preserving solution for intelligent systems in resource-constrained environments. The main contributions of this paper are summarized as follows.

(1) We present a theoretical interpretation of the essence of intelligence, identifying its core as the capability to employ knowledge for problem solving and task completion.

(2) We introduce the first principle of intelligence that models a goal-driven process to accomplish tasks.

(3) We develop a collaborative intelligence paradigm driven by knowledge, termed the six-capability network.

(4) We design an networked intelligence system architecture grounded in the six-capability network, enabling knowledge-level cooperation among small model-based agents.

II. The Essence of Intelligence

Intelligence, at its core, is the capability of an agent to utilize knowledge to solve problems and accomplish tasks under environmental constraints. Rather than being a single skill or isolated function, intelligence is a composite ability emerging from the coordinated operation of multiple cognitive processes. From a functional perspective, it reflects the integrated capacity of a system to perceive, interpret, learn, adapt, and act purposefully toward goals[23-25].

Formally, consider an agent interacting with an environment characterized by a state space S, an action space A, and a task objective G. Let $s_t \in S$ denote the environment state at time t, and let K represent the agent's knowledge base, including prior experience and learned representations. Intelligence can be expressed as the agent's ability to construct a strategy π such that

$$\pi : (s_t, K, G) \to a_t, \quad s_t \in S,\ a_t \in A, \tag{1}$$

where actions are inferred through knowledge to move the system toward task completion.

A defining characteristic of intelligence is that it gradually reduces the uncertainty of information. Given an initial state s_0 with high uncertainty, the agent transforms the environment through a sequence of actions a_t into a target state s_t aligned with the goal G. This process can be viewed as a reduction in entropy[26-27].

$$H(S_t \mid K_t) < H(S_0 \mid K_0), \tag{2}$$

where $H(\cdot)$ denotes information entropy, S_t denotes the random variable of system states at time t, and K_t denotes the knowledge available at that time. Intelligence thus operates as a mechanism that compresses uncertainty through structured perception, inference, and action.

To realize this capability, intelligence manifests through several tightly coupled functional components[28-29]. Firstly, perception and understanding enable the agent to map raw sensory inputs to meaningful internal representations. In humans, this corresponds to visual, auditory, and linguistic comprehension; in artificial systems, it includes image, speech, and text understanding. The essential outcome is semantic representation rather than mere signal detection.

Secondly, learning and memory allow the agent to extract regularities from experience and retain useful information over time. Learning updates the knowledge base K, while memory ensures that past experience can inform future decisions. Without learning, intelligence would be confined to fixed-rule behavior and unable to improve or generalize.

Thirdly, reasoning and decision-making operate on perceived and remembered information to infer relationships, evaluate alternatives, and select actions. Reasoning transforms knowledge into actionable insight, while decision-making resolves trade-offs under constraints. These processes enable the agent to move beyond proactive behavior toward deliberative control.

Fourthly, adaptation and innovation distinguish intelligence from mechanical repetition. When encountering novel states $s \notin S_{\text{seen}}$, an agent modifies its strategy π rather than rigidly applying past solutions. In advanced cases, this adaptability enables the generation of new strategies or representations, reflecting creative problem-solving.

Finally, goal orientation unifies all components of intelligence. Given an objective G, the agent actively coordinates perception, memory, reasoning, and action to transform the environment toward the desired outcome. Intelligence is therefore inherently purposive, not merely reactive.

From this perspective, knowledge constitutes the core soft power of intelligence. Knowledge encodes priors that guide inference and drastically reduce the need for brute-force exploration. Systems lacking knowledge must compensate with excessive computation or energy, whereas intelligent systems can act efficiently to empower various industries.

From an information-theoretic standpoint, intelligence reduces the effective entropy of the environment[30]. Richer perception and deeper cognition lead to more informative state representations, which in turn lower uncertainty. Importantly, data entropy is inversely related to prior knowledge: stronger priors yield more compact representations and lower processing costs. Consequently, intelligence transforms high-entropy, difficult-to-control systems into low-entropy, manageable ones.

In summary, intelligence can be understood as a goal-directed, knowledge-driven, closed-loop process that reduces uncertainty through perception, reasoning, action, and learning. This understanding provides the theoretical basis for rethinking AI system architectures beyond purely statistical models.

III. The First Principle of Intelligence

Building upon the essence of intelligence discussed in section II, this section proposes the formulation of the first principle of intelligence, which models intelligence as a goal-oriented cognitive process that transforms raw physical signals into actionable strategies.

A. From Physical States to Intelligent Action

Intelligence begins with the interaction between an agent and the physical world. Let w_t denote the physical state of the environment at time t. This physical state generates observable signals (e.g., light, sound, electromagnetic waves), which are captured by sensors and transformed into data. Through successive processing stages, raw data is converted into semantic representations and, ultimately, into executable strategies.

This transformation can be abstracted as

$$w_t \rightarrow x_t \rightarrow s_t^p \rightarrow s_t^c \rightarrow \pi_t, \tag{3}$$

where x_t denotes sensory data, s_t^p denotes perceptual semantics, s_t^c denotes cognitive semantics, and π_t denotes a strategy or policy. This sequence highlights that intelligence is not derived directly from data, but from structured semantic processing that bridges perception and action.

The execution of the strategy π_t produces an action a_t, which alters the physical state of the environment

$$w_{t+1} = f(w_t, a_t). \tag{4}$$

The intelligence of the system is reflected in its ability to select a_t such that the resulting state w_{t+1} moves closer to a desired goal.

B. Functional Modules of Intelligence

Based on this transformation pipeline, the first principle of intelligence can be decomposed into a set of essential functional modules. (1) Perceiving information: mapping sensory data to perceptual semantics. (2) Understanding information: transforming perceptual semantics into cognitive semantics. (3) Discovering knowledge: extracting patterns and regularities to update the knowledge base. (4) Using knowledge: applying prior knowledge to interpret states and constrain decisions. (5) Creating knowledge: generating new abstractions or hypotheses when existing knowledge is insufficient. (6) Expressing strategies: converting cognitive decisions into executable plans. (7) Executing strategies: applying actions to the environment to change its state.

Let K_t denote the knowledge base at time t, and let s_t denote the estimated cognitive state. The intelligent process can be summarized as

$$(s_t, K_t) \xrightarrow{\text{reasoning}} \pi_t \xrightarrow{\text{execution}} s_{t+1}, \quad K_{t+1} = K_t \cup \Delta K_t, \tag{5}$$

where π_t denotes the inferred strategy, and ΔK_t represents newly acquired knowledge.

C. The First Principle

The first principle of intelligence can now be stated as follows.

> Intelligence is a goal-oriented, closed-loop process in which an agent, under external constraints, leverages knowledge to evaluate discrepancies between its internal state and task objectives, formulates strategies to reduce these discrepancies, and iteratively executes actions through the coordinated operation of multiple functional processes to achieve goal completion.

This principle implies that no single functional module is sufficient to realize intelligence. Instead, intelligence emerges from their interaction and mutual constraint. Formally, let $\{M_1, M_2, \cdots, M_n\}$ denote the set of functional modules. Intelligence I is not a linear sum of these modules, but a coordinated composition, defined as

$$I \neq \sum_{i=1}^{n} M_i, \quad I = \mathscr{C}(M_1, M_2, \cdots, M_n), \tag{6}$$

where $\mathscr{C}(\cdot)$ denotes a coordination operator encoding information flow, consistency constraints, and feedback loops.

A critical implication of this formulation is that intelligence cannot be realized by a single monolithic process. The functional processes described above operate on different representations, time scales, and objectives. Attempting to compress perception, reasoning, memory, and action into a single giant model inevitably leads to poor interpretability, limited controllability, and excessive resource consumption.

Therefore, the first principle naturally leads to collaborative intelligence, where intelligent behavior emerges from the cooperation of multiple small, specialized models. Let $\{f_1, f_2, \cdots, f_n\}$ denote a set of such models, each responsible for a specific functional process. Intelligent behavior arises through their coordinated interaction, expressed as

$$\pi_t = f_n \circ f_{n-1} \circ \cdots \circ f_1(x_t). \tag{7}$$

Collaboration is not a simple feed-forward pipeline. Each model produces intermediate representations that must be aligned semantically, which can be understood as consensus intelligence. This paper formulate consensus intelligence under the theoretical framework of "semantic in-position" and "semantic off-position"[31]. A semantic in-position signal refers to the standardized semantic feature representation conventionally associated with a symbolic identity. For example, every Chinese character has a standard pronunciation, which can be called semantic in-position. A semantic off-position signal refers to a parameter-shifted semantic feature expression derived from the in-position signal, whose attribute values differ from the canonical configuration but whose symbolic identity remains invariant. For example, a Chinese character may be pronounced differently across various regional dialects, yet they are still identifiable as the same character. Such pronunciations constitute semantic off-position signals. Let $\mathscr{Z}$ denote the unified semantic representation space. For a given physical state s, we denote its corresponding in-position semantic signal by $z \in \mathscr{Z}$. Each module f_i produces a semantic output $z_i = f_i(x_t) \in \mathscr{Z}$. We measure the deviation of a module output from the semantic in-position prototype as

$$\delta_i = d(z_i, z), \tag{8}$$

where $d : \mathscr{Z} \times \mathscr{Z} \to \mathbb{R}_{\geqslant 0}$ denotes the semantic distance in the unified semantic space. The confidence contributed by module i to a candidate in-position signal z is defined as

$$\gamma_i(z) = \exp\left(-\frac{1}{2\sigma_i^2} d^2(z_i, z)\right), \tag{9}$$

where σ_i characterizes the intrinsic uncertainty of module f_i. A smaller deviation implies higher semantic confidence. The desired property of the six-capability network is that, for the same physical state, all module outputs remain consistent with the same semantic in-position signal. The consensus semantic representation is therefore obtained by maximizing the joint semantic confidence, expressed as

$$z_t^* = \arg\max_{z \in \mathscr{Z}} \prod_{i=1}^{n} \gamma_i(z). \tag{10}$$

Small models are the natural carriers of collaborative intelligence. They are modular, interpretable, and replaceable; they can be tuned independently according to task requirements; and they operate under significantly lower computational and energy constraints. Moreover, small models can be organized into networked structures rather than isolated systems.

Let a collaboration network be represented as a graph $G = (V, E)$, where each node corresponds to a small model and edges represent semantic communication. The internal state of each model evolves as

$$z_v^{t+1} = \Phi_v\left(z_v^t, \{z_u^t \mid (u, v) \in E\}\right). \tag{11}$$

Through continuous interaction, the network converges toward coherent, goal-aligned behavior. In this view, intelligence is not centralized but distributed, emerging from coordination rather than scale.

IV. A Knowledge-Driven Collaborative Architecture: the Six-Capability Network

Building upon the first principles of intelligence discussed in the previous section, where intelligence is regarded as the

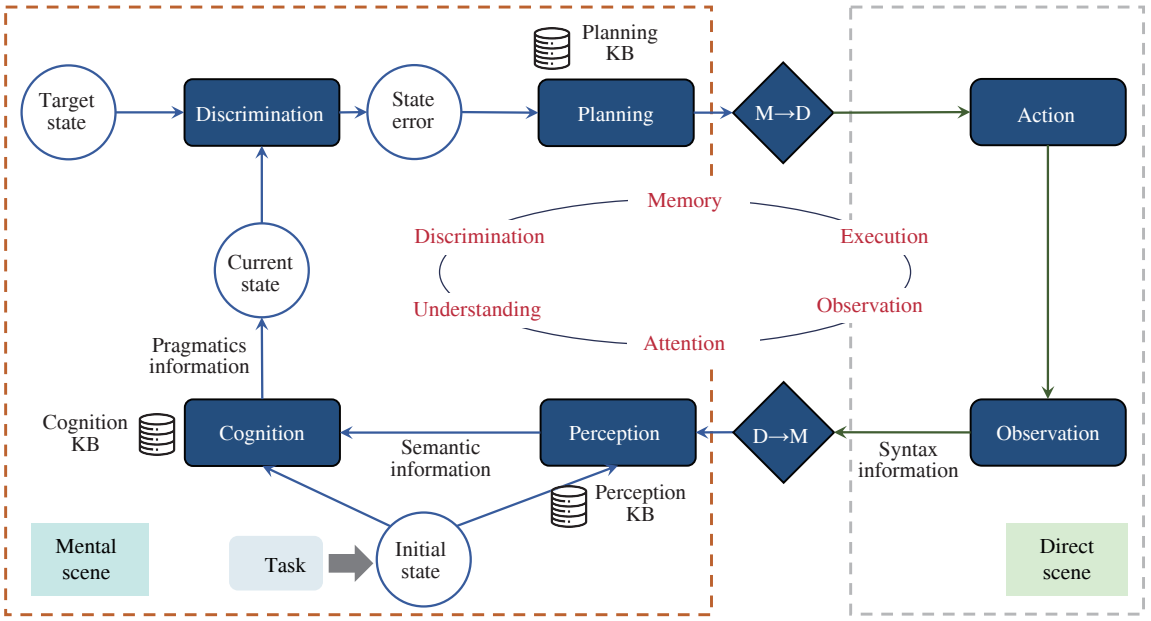

Fig. 2 The framework of six-capability network

result of coordinated interactions among multiple functional capabilities rather than a monolithic system, this section introduces a knowledge-driven collaborative architecture, referred to as the "six-capability network". The core idea is that intelligence emerges from the structured cooperation of specialized capabilities, organized around task-driven state transitions and closed-loop control.

Different from data-driven large-model paradigms, the six-capability network is explicitly designed to be knowledge-driven and small-model-based. Small models refer to lightweight, modular functional models designed to implement a single cognitive capability, rather than compressed variants of large foundation models. When a task exceeds the local knowledge scope, the system does not transmit raw sensory data or user information. Instead, it selectively queries cloud-based large models for missing or abstract knowledge, ensuring that what circulates within the network is knowledge rather than data, thereby fundamentally preserving user and institutional privacy.

The six-capability network takes biological intelligence as its structural and functional reference[17] and proposes a six-dimensional capability space, consisting of observation, attention, discrimination, understanding, memory, and execution. Instead of emphasizing static input-output mappings, the proposed architecture focuses on modeling how an intelligent system drives the environment from an initial state toward a target state under task constraints, forming a knowledge driven closed-loop process. When a task exceeds the local knowledge scope, the system does not transmit raw sensory data or user information. Instead, it selectively queries cloud-based large models for missing or abstract knowledge, ensuring that what circulates within the network is knowledge rather than data, thereby fundamentally preserving user and institutional privacy. The framework is shown in Fig. 2.

Formally, intelligence is described in a state-space framework. Let

$$S = \{s_i \mid s_i \text{ denotes the system state at a given time}\}, \quad (12)$$

represent the set of all possible states of the environment and the agent. A state is a symbolic abstraction of the physical or semantic world, composed of task-relevant variables, such as position, configuration, semantic labels, or logical relations. The evolution of the system is governed by a state transition function

$$T : S \times A \to S, \quad (13)$$

where A denotes the set of executable actions. Given a current state $s \in S$ and an action $a \in A$, the system transitions to a new state $s' = T(s,a)$.

Under this formulation, an intelligent task can be represented as a sequence of state transitions

$$s_0 \xrightarrow{a_1} s_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} s_n = G, \quad (14)$$

where s_0 is the initial state when the task is issued, and G is the target state representing task completion. G originates from the user's task intention. The intention is semantically parsed into a structured representation of desired properties. For example, the intention "clean the room" can be translated into formalized semantic constraints such as "floor state = clean". The essence of intelligence lies in inferring an effective sequence of actions $\{a_1, \cdots, a_n\}$ that progressively reduces uncertainty and transforms the system state toward the goal.

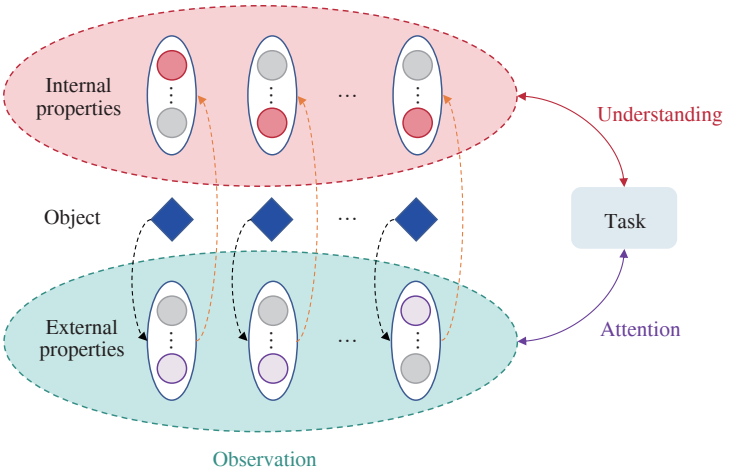

Fig. 3 Diagram of observation, attention, and understanding in six-capability network

Then, this paper introduces the degree of intelligence (D_{AI}) to evaluate the capability of an intelligent system from two aspects—transition cost and transition accuracy

$$D_{\mathrm{AI}} = F(c_{\mathrm{trans}}, e_{\mathrm{trans}}), \tag{15}$$

where $c_{\mathrm{trans}} = g(t,S,E)$ denotes the transition cost, computed from time consumption t, spatial resource usage S, and energy expenditure E, and $e_{\mathrm{trans}} = \|f_0 - f_1\|$ denotes the transition error, defined as the deviation between the current state and the target state. Lower transition cost and smaller state deviation correspond to a higher degree of intelligence.

Within the transition process, the six capabilities of the proposed model assume complementary roles. Observation serves as the grounding and feedback mechanism of the system, transforming inputs from the direct scene (the real-world environment) into the mental scene (the internally imagined scenario of the agent). It continuously acquires environmental signals E and maps them into symbolic state representations, capturing observable external properties (e.g., color, shape), as illustrated by the dashed ellipse in Fig. 3.

$$E \xrightarrow{\text{Observation}} s_i. \tag{16}$$

This capability is implemented using lightweight convolutional neural networks (CNNs), whose trained parameters are perceptual knowledge.

Attention functions as an information-filtering mechanism that transforms the symbolic outputs of observation into semantic vector representations and computes their relevance to the task (shown as purple circles in Fig. 3). $s_i = \{a_k\}_{k=1}^{K}$ represents a set of semantic properties from observation, attention computes the task-dependent relevance of each property with respect to the task representation t. The relevance weight of attribute a_k is defined as

$$\alpha_k = \frac{\exp(\mathrm{sim}(a_k,t))}{\sum_{j=1}^{K}\exp(\mathrm{sim}(a_j,t))}, \tag{17}$$

where $\mathrm{sim}(\cdot,\cdot)$ measures semantic correlation. The task-focused external properties is then obtained through weighted aggregation, expressed as

$$s_i^p = \sum_{k=1}^{K} \alpha_k a_k. \tag{18}$$

Understanding models internal properties of objects through structured and hierarchical symbolic reasoning over an property knowledge base. It compresses semantic information into task-relevant representations while preserving reconstructability, generating pragmatics information that support reasoning and planning, expressed as

$$s_i^p \xrightarrow{\text{Understanding}} s_i^c, \tag{19}$$

where s_i^c infers task-relevant internal properties based on external properties from attention (shown as red circles above in Fig. 3). This process is grounded in a cognitive knowledge base, which is a structured symbolic repository of object attributes. It is primarily constructed based on expert knowledge and domain-specific rules.

Discrimination evaluates the discrepancy between the current state and the target state. The current state $s_i^c = \{a_k^{\mathrm{current}}\}_{k=1}^{N}$, where each a_k^{current} denotes a property. The target state $G = \{a_k^{\mathrm{target}}\}_{k=1}^{N}$, where each a_k^{target} represents the desired value of the corresponding property. This discrepancy can be expressed as a state difference

$$\Delta s_i = \frac{1}{N}\sum_{k=1}^{N} \left\| a_k^{\mathrm{target}} - a_k^{\mathrm{current}} \right\|^2. \tag{20}$$

If Δs_i cannot be reduced after multiple decision iterations, the system infers that the local decision knowledge base lacks the necessary knowledge to accomplish the current subtask. In such cases, the subtask is abstracted into a high-level se-

mantic description, and only non-sensitive symbolic information is transmitted to a cloud-based large model for knowledge supplementation. The cloud model returns abstract rules or structured knowledge, which are then incorporated into the local decision knowledge base for memory module.

Memory maintains a structured and hierarchical knowledge base describing state transfer laws and derives transfer strategies based on the computed state distance. It maintains historical state-action-state patterns, expressed as

$$\mathscr{M} = \{(s, a, s^p)\}, \tag{21}$$

and retrieves relevant experiences to support action selection and strategy generation, enabling generalization and transfer across tasks.

Execution translates cognitive decisions into concrete actions, grounding them in the direct scene from the mental scene. Given an action a_i, execution invokes the state transition function to advance the system

$$s_{i+1} = T(s_i, a_i). \tag{22}$$

Execution grounds intelligence in physical or virtual interaction with the environment.

Together, these capabilities form a dynamic closed-loop process, expressed as

$$E \to s_0 \to s_0^p \to s_0^c \to \Delta s_0 \to a_1 \to s_1 \to \cdots \to G, \tag{23}$$

where observation initiates and validates each iteration, and the loop continues until the target state is reached. Crucially, this closed loop operates primarily on internal state abstractions and knowledge representations, allowing intelligent behavior to emerge without exposing raw environmental data beyond the local system boundary.

To ensure efficiency across tasks of varying complexity, the six-capability network incorporates a task-adaptive activation mechanism. Let

$$C \subseteq \{\text{Observation}, \text{Attention}, \text{Understanding}, \text{Discrimination}, \text{Memory}, \text{Execution}\}, \tag{24}$$

denote the set of activated capabilities for a given task. Simple tasks activate minimal processing paths, while complex, long-horizon tasks dynamically activate the full closed loop. This on-demand activation strategy enables energy-efficient operation, supports edge deployment, and provides a scalable foundation for collaborative intelligence based on coordinated small models rather than a single large model.

V. Knowledge-Driven Intelligence vs LLM-Driven Intelligence

From the perspective of intelligence paradigms, the knowledge-driven six-capability network and LLM-driven AI agents represent two fundamentally different approaches to constructing autonomous intelligent systems. From a paradigmatic perspective, although both frameworks share the superficial view that intelligence can be abstractly formulated as a mapping from observations to actions under task constraints, they differ fundamentally in their structural assumptions.

Firstly, the two paradigms differ in the origin and organization of intelligence. In LLM-driven AI agents, intelligence is primarily encoded in a large-scale parameter space shaped through exposure to massive datasets. Let $\boldsymbol{\theta}$ denote the model parameter vector, then an agent's behavior can be represented as

$$a = f_{\boldsymbol{\theta}}(s), \tag{25}$$

where knowledge, reasoning heuristics, and behavioral tendencies are all implicitly entangled within $\boldsymbol{\theta}$. This entanglement makes it difficult to isolate or directly modify specific forms of knowledge. By contrast, the six-capability network adopts a knowledge-centered organizational structure, in which intelligence arises from explicit knowledge interacting with structured capability modules. Action generation follows the form

$$a = g(s, K), \tag{26}$$

where K denotes a transparent and extensible knowledge base. Knowledge thus becomes an independent and manipulable entity, rather than being absorbed into parameters.

Secondly, the two approaches employ distinct learning mechanisms. LLM-driven systems learn primarily through large-scale statistical optimization

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathbb{E}_{(x,y)\sim D}\left[\mathscr{L}(f_{\boldsymbol{\theta}}(x), y)\right], \tag{27}$$

with intelligence emerging as a byproduct of distribution fitting. The six-capability network, however, treats learning as a process of accumulating and refining knowledge, which can be represented abstractly as

$$K_{t+1} = K_t \cup \{(s, a, s')\}. \tag{28}$$

Here, learning modifies explicit symbolic or structured knowledge, making the evolution of intelligence interpretable and traceable.

Thirdly, the paradigms differ in their reasoning processes. In LLM-driven agents, reasoning is an emergent property of high-dimensional parameter interactions, lacking guaranteed semantic correspondence between intermediate internal states and meaningful reasoning steps. In contrast, reasoning within the six-capability network is knowledge-guided and module-oriented, proceeding through distinct capability stages, such as state understanding, goal discrimination, and action selection. Each stage may be inspected, modified, or recomposed, supporting interpretability and controllability.

Fourthly, the two paradigms diverge in task customization. LLM-driven agents embed multiple tasks within a shared parameter space, making structural customization difficult without retraining or fine-tuning. The six-capability network, however, enables task-adaptive activation of capability modules, allowing different configurations to be selectively engaged while preserving computational efficiency.

A deeper contrast emerges when intelligence is examined from a geometric perspective. Task execution can be regarded as finding a feasible trajectory in a high-dimensional state space from an initial state s_0 to a goal G

$$\gamma : s_0 \to G. \tag{29}$$

In LLM-driven agents, this nonlinear trajectory is implicitly approximated through layered parameterized transformations, which may be viewed as a form of large-scale piecewise linear approximation

$$\gamma \approx \sum_i W_i s + b_i. \tag{30}$$

The resulting trajectory is encoded implicitly within the parameters, making the mapping between parameter updates and reasoning behavior neither direct nor interpretable.

By contrast, in the six-capability network, the trajectory is realized through explicit state transitions guided by knowledge

$$s_{i+1} = T(s_i, a_i), \;\; a_i = g(s_i, K). \tag{31}$$

Each segment of the trajectory corresponds to a meaningful decision grounded in understanding, knowledge, and environmental context. Thus, the reasoning path becomes observable, explainable, and controllable at the level of both knowledge and capability interactions.

In summary, LLM-driven AI agents pursue intelligence primarily through large-scale parameter optimization and statistical correlation learning, whereas the six-capability network pursues intelligence through structured knowledge representation, modular capability coordination, and explicit state-space reasoning. The former emphasizes data and scale; the latter emphasizes understanding, control, and task-oriented adaptability. Viewed together from both paradigm and geometric perspectives, the six-capability network offers a complementary and potentially more sustainable direction for building interpretable, customizable, and resource-efficient artificial intelligence systems.

VI. Networked Intelligence Systems Based on the Six-Capability Network

Building upon the paradigm of knowledge-driven intelligence, we extend intelligence from isolated agents to networked systems. If intelligence emerges from the coordination of multiple capabilities within a single agent, then large-scale intelligent behavior naturally arises from the collaboration of multiple such agents connected through a structured communication mechanism. This motivates a networked intelligence system based on the six-capability network.

In the proposed system, intelligence is not centralized in a single large model, nor distributed as independent black-box agents. Instead, it is realized through a semantic communication network that connects multiple intelligent agents, each instantiated as a small-model system grounded in the six-capability architecture and equipped with its own knowledge base. Each agent specializes in certain functions or domains, and global intelligence emerges through their coordinated interaction. When a task requirement exceeds the agent's internal knowledge scope, it actively requests relevant knowledge from other agents through semantic communication, and global intelligence emerges from this capability-level knowledge collaboration. The system is shown in Fig. 4.

Formally, let

$$\mathscr{A} = \{A_1, A_2, \cdots, A_N\}, \tag{32}$$

denote a set of intelligent agents deployed within the system. Each agent A_i is defined as

$$A_i = (C_i, K_i), \tag{33}$$

where C_i represents an instance of the six-capability network with task-adaptive activation, and K_i denotes the agent's local knowledge base. The knowledge base may contain factual knowledge, procedural rules, domain constraints, or learned state-transition patterns, depending on the agent's role. Semantic consistency is ensured by grounding all knowledge in a unified semantic space, where every agent is built upon a shared structured framework with consistent symbolic representations and encoding rules from the outset.

Unlike conventional multi-agent systems that exchange raw data, intermediate representations, or parameter updates, the proposed architecture adopts a semantic communication paradigm. Communication between agents is conducted through semantic symbols rather than data streams. Let

$$\Sigma = \{\sigma_1, \sigma_2, \cdots\}, \tag{34}$$

denote the space of semantic symbols. A transmitted message $\sigma \in \Sigma$ represents structured meaning, such as task intent, state abstraction, causal relations, constraints, or action suggestions. Fundamentally, a semantic symbol encodes knowledge rather than raw data, serving as a compact and interpretable carrier of problem-relevant information.

Thus, inter-agent communication can be expressed as

$$A_i \xrightarrow{\sigma} A_j, \tag{35}$$

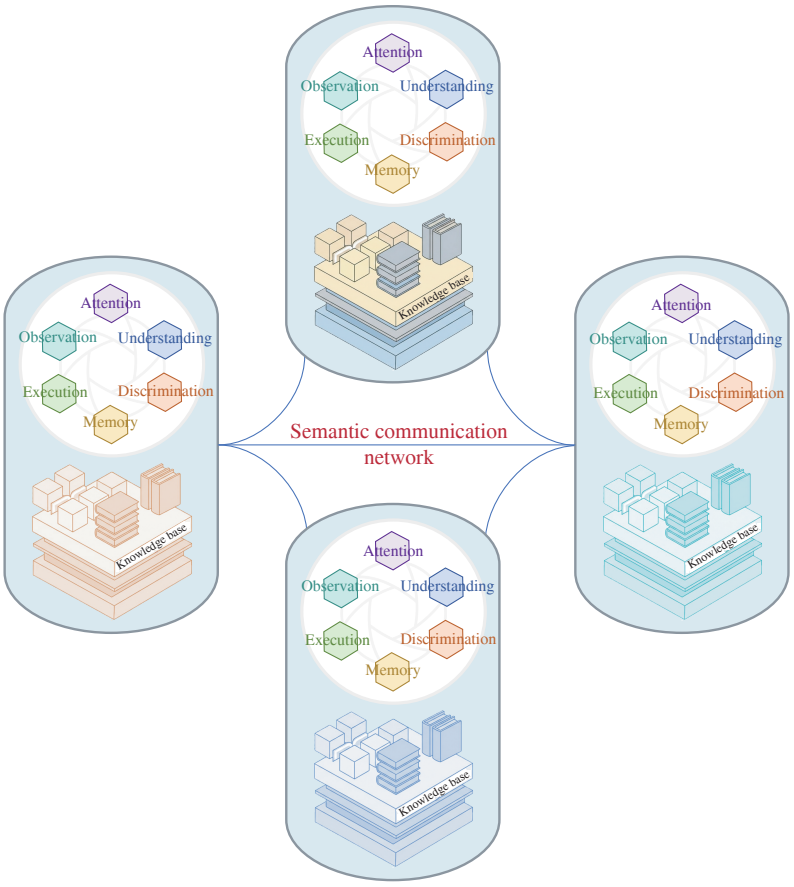

Fig. 4 Networked intelligence systems based on the Six-Capability Network

where σ is directly interpretable by the receiving agent and can be integrated into its understanding, discrimination, or memory modules without requiring large-scale decoding or statistical inference. This mechanism enables the system to transmit knowledge rather than data, significantly reducing communication bandwidth and computational overhead.

Within this semantic communication network, agents collaboratively solve tasks by sharing partial understandings and complementary knowledge. From a system perspective, task execution corresponds to a distributed state-space transition process. Let s_i^t denote the internal state of agent A_i at time t. The global system state can be abstractly represented as

$$S^t = \{s_1^t, s_2^t, \cdots, s_N^t\}. \tag{36}$$

Through semantic exchanges and local reasoning, the system evolves toward a task goal state G, even though no single agent necessarily holds a complete global model. Importantly, collaboration is achieved by sharing partial knowledge and semantic abstractions, rather than by exposing local data or internal states.

An important property of the proposed system is the modularity and extensibility of knowledge. Knowledge bases K_i are independent, incrementally expandable, and replaceable. New agents can be added to the network without retraining existing ones, and outdated knowledge can be revised locally. To ensure consistency, traceability, and trustworthiness of shared knowledge, the system may incorporate a blockchain-based ledger to record knowledge updates, provenance, and access permissions. In this way, knowledge sharing becomes auditable and decentralized, avoiding single points of failure.

From an architectural perspective, the proposed networked intelligence system is a distributed collaborative paradigm. Each six-capability network constitutes a self-contained intelligent unit equipped with an integrated capability structure that enables independent perception, attention, understanding, discrimination, memory, and execution, thereby forming a full cognitive closed loop. Collaboration among multiple six-capability networks is not a simple aggregation of fragmented functions, but a coordination among autonomous agents that each possess fundamental intelligence capabilities. This principle is structurally distinct from conventional distributed AI architectures in which individual nodes typically perform partial functions and depend on collective system integration to achieve intelligence. In the proposed framework, functional roles are explicitly decomposed, state representations are clearly structured, and coordination is conducted

through explicit state-space transitions and discrepancy computations, ensuring strong interpretability and cognitive-level transparency.

VII. Conclusion

This paper systematically reexamines the foundations of intelligence from both practical demands and theoretical principles. We first analyze the inherent limitations of the large-model paradigm and show that its data- and scale-centric nature cannot fundamentally satisfy the growing demand for intelligent systems that are private, locally deployable, and resource-efficient. To narrow this gap, we clarify the essence of intelligence as the capability to leverage knowledge to solve problems and accomplish tasks under constraints, rather than as a byproduct of model scale or data accumulation. Based on this understanding, we formulate the first principle of intelligence, modeling intelligence as a closed-loop, goal-oriented process that continuously reduces the discrepancy between internal states and task objectives through knowledge-guided reasoning and action. Guided by this principle, we introduce the six-capability network as a modular, knowledge-driven cognitive architecture, and further extend it to a networked intelligence system in which multiple agents collaborate through semantic-level communication. Together, these contributions outline a coherent alternative intelligence paradigm that decouples intelligence from scale and provides a viable path toward interpretable, decentralized, and deployable intelligent systems.

Nevertheless, the limitation should be acknowledged. The proposed framework relies on structured knowledge construction, which requires expert involvement during the knowledge modeling stage. Its primary strength lies in knowledge utilization and dynamic scheduling rather than in the autonomous discovery of fundamentally new scientific laws. The system is not intended to create new scientific knowledge. In highly open domains where knowledge structures remain unclear or incomplete, symbolic reasoning may encounter scalability and generalization challenges. The framework is particularly advantageous in privacy-sensitive, resource-constrained, and explainability-critical scenarios, such as home environments, industrial manufacturing, defense systems, educational institutions, and edge computing systems.

Conflict of interest statement. None declared.

References

[1] Bommasani R. On the opportunities and risks of foundation models[J]. arXiv preprint arXiv:2108.07258, 2021.

[2] Eder M, Sjøvaag H. Artificial intelligence and the dawn of an algorithmic divide[J]. Frontiers in Communication, 2024, 9: 1453251.

[3] Somvanshi S, Islam M M, Chhetri G, et al. From tiny machine learning to tiny deep learning: a survey[J]. ACM Computing Surveys, 2025, 58(7): 1-33.

[4] Wu Y, Sun Z, Li S, et al. Inference scaling laws: an empirical analysis of compute-optimal inference for LLM problem-solving[J]. arXiv preprint arXiv:2408.00724, 2024.

[5] Thompson N C, Greenewald K, Lee K, et al. The computational limits of deep learning[J]. arXiv preprint arXiv:2007.05558, 2020.

[6] Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP[J]. arXiv preprint arXiv:1906.02243, 2019.

[7] Patterson D, Gonzalez J, Le Q, et al. Carbon emissions and large neural network training[J]. arXiv preprint arXiv:2104.10350, 2021.

[8] Lacoste A, Luccioni A, Schmidt V, et al. Quantifying the carbon emissions of machine learning[J]. arXiv preprint arXiv:1910.09700, 2019.

[9] Schwartz R, Dodge J, Smith N A, et al. Green AI[J]. Communications of the ACM, 2020, 63(12): 54-63.

[10] OECD. Measuring the environmental impacts of AI compute and applications: the AI footprint[R]. Paris: OECD Publishing, 2022.

[11] Shumailov I, Shumaylov Z, Zhao Y, et al. AI models collapse when trained on recursively generated data[J]. Nature, 2024, 631(8022): 755-759.

[12] Maslej N, Fattorini L, Perrault R, et al. Artificial intelligence index report 2025[R]. 2025.

[13] Zhuang B, Liu J, Pan Z, et al. A survey on efficient training of transformers[C]//Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. New York: ACM Press, 2023: 6823-6831.

[14] Reed S, Zolna K, Parisotto E, et al. A generalist agent[J]. arXiv preprint arXiv:2205.06175, 2022.

[15] Mumuni F, Mumuni A. Explainable artificial intelligence (XAI): from inherent explainability to large language models[J]. arXiv preprint arXiv:2501.09967, 2025.

[16] Malwade S S, Budhavale S J. Exploring explainable AI: current trends, challenges, techniques and its applications[C]//Proceedings of the 5th International Conference on Information Management and Machine Intelligence. New York: ACM Press, 2023: 1-9.

[17] Castro-Alonso J C, Hidalgo A A, Sweller J. Biological evolution and human cognition are analogous information processing systems[J]. Frontiers in Psychology, 2024, 14: 1330345.

[18] Gou J, Yu B, Maybank S J, et al. Knowledge distillation: a survey[J]. International Journal of Computer Vision, 2021, 129(6): 1789-1819.

[19] Zhu X, Li J, Liu Y, et al. A survey on model compression for large language models[J]. Transactions of the Association for Computational Linguistics, 2024, 12: 1556-1577.

[20] Hou X, Han J, Zhao Y, et al. Unveiling the landscape of LLM deployment in the wild: an empirical study[J]. arXiv preprint arXiv:2505.02502, 2025.

[21] Dodig-Crnkovic G. Rethinking cognition: morphological info-computation and the embodied paradigm in life and artificial intelligence[M]//Scientific Cognition, Semiotics, and Computational Agents: Essays in Honor of Lorenzo Magnani-Volume 2. Cham: Springer Nature Switzerland, 2025: 65-87.

[22] Red'ko V G, Samsonovich A V, Klimov V V. Computational modeling of insight processes and artificial cognitive ontogeny[J]. Cognitive Systems Research, 2023, 78: 71-86..

[23] Lake B M, Ullman T D, Tenenbaum J B, et al. Building machines that learn and think like people[J]. Behavioral and Brain Sciences, 2017, 40: 253.

[24] LeCun Y. A path towards autonomous machine intelligence[J]. Open Review, 2022, 62(1): 1-62.

[25] Marcus G. The next decade in AI: four steps towards robust artificial intelligence[J]. arXiv preprint arXiv:2002.06177, 2020.

[26] Friston K. The free-energy principle: a unified brain theory?[J]. Nature Reviews Neuroscience, 2010, 11(2): 127-138.
[27] Tishby N, Zaslavsky N. Deep learning and the information bottleneck principle[C]//Proceedings of 2015 IEEE Information Theory Workshop (ITW). Piscataway: IEEE Press, 2015: 1-5.
[28] Scholkopf B, Locatello F, Bauer S, et al. Toward causal representation learning[J]. Proceedings of the IEEE, 2021, 109(5): 612-634.
[29] Goyal A, Bengio Y. Inductive biases for deep learning of higher-level cognition[J]. Proceedings of the Royal Society A, 2022, 478(2266): 20210068.
[30] Polyanskiy Y, Wu Y. Information theory: from coding to learning[M]. Cambridge: Cambridge University Press, 2025.
[31] Shi G M, Gao D H. Mathematical characterization for semantic information and its signal content[J]. Scientia Sinica Informationis, 2025, 55(8): 1950-1970.

About the Authors



Guangming Shi (Fellow, IEEE) [corresponding author] received the B.S. degree in automatic control, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, Xi'an, China, in 1985, 1988, and 2002, respectively. From 1994 to 1996, he was a Research Assistant cooperated with the Department of Electronic Engineering, The University of Hong Kong, Hong Kong, China. He joined the School of Electronic Engineering, Xidian University, in 1988. Since 2003, he has been a Professor with the School of Electronic Engineering, Xidian University, Xi'an, China. In 2004, he was the Head of the National Instruction Base of Electrician and Electronic (NIBEE), Xi'an, China. From June 2004 to December 2004, he studied with the Department of Electronic Engineering, University of Illinois at Urbana-Champaign (UIUC), Champaign, USA. He is currently the Deputy Director of the School of Electronic Engineering, Xidian University, and the Academic Leader in the subject of circuits and systems. He has authored or coauthored more than 100 research articles. His research interests include compressed sensing, the theory and design of multirate filter banks, image denoising, low-bit-rate image/video coding, and the implementation of algorithms for intelligent signal processing (using DSP and FPGA).



Ning Lan received the M.S. degree from Shanghai University of Engineering Science, Shanghai, China, in 2019. She is currently pursuing the Ph.D. degree at Xidian University, Xi'an, China. Her current research interests include embodied intelligence, large language model, and computer vision.



Xuemei Xie (Senior Member, IEEE) received the M.S. degree in electronic engineering from Xidian University, Xi'an, China, in 1994, and the Ph.D. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, China, in 2004. She is currently a Professor with the School of Artificial Intelligence, Xidian University, Xi'an, China. She has authored more than 50 academic papers in international and national journals, and international conferences. Her research interests include artificial intelligence, compressive sensing, deep learning, image and video processing, and filter banks.



Jinjian Wu received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2008 and 2014, respectively. He is currently a Professor and Ph.D. Supervisor with Xidian University, where he was promoted to Professor in 2019 with exceptional qualifications. He is a recipient of several prestigious young talent awards in China, including the National Young Talent Program and the Huo Yingdong Young Faculty Foundation. His research interests include high-quality imaging and intelligent image processing, such as event-based camera imaging systems, image quality enhancement and assessment, and object detection and recognition. He has authored or coauthored over 100 papers, including more than 50 recent publications in top journals and conferences. He has led multiple national-level research projects and received the Second Prize of the National Natural Science Award of China and the First Prize of the Natural Science Award of Shaanxi Province.