

# Intelligence Foundation Model: A New Perspective to Approach Artificial General Intelligence

Borui Cai, Yao Zhao

**Abstract**—We propose a new perspective for approaching artificial general intelligence (AGI) through an intelligence foundation model (IFM). Unlike existing foundation models (FMs), which specialize in pattern learning within specific domains such as language, vision, or time series, IFM aims to acquire the underlying mechanisms of intelligence by learning directly from diverse intelligent behaviors. Vision, language, and other cognitive abilities are manifestations of intelligent behavior; learning from this broad range of behaviors enables the system to internalize the general principles of intelligence. Based on the fact that intelligent behaviors emerge from the collective dynamics of biological neural systems, IFM consists of two core components: a novel network architecture, termed the state neural network, which captures neuron-like dynamic processes, and a new learning objective, neuron output prediction, which trains the system to predict neuronal outputs from collective dynamics. The state neural network emulates the temporal dynamics of biological neurons, allowing the system to store, integrate, and process information over time, while the neuron output prediction objective provides a unified computational principle for learning these structural dynamics from intelligent behaviors. Together, these innovations establish a biologically grounded and computationally scalable foundation for building systems capable of generalization, reasoning, and adaptive learning across domains, representing a step toward truly AGI.

**Index Terms**—Artificial general intelligence, Foundation model, Deep neural networks.

## I. INTRODUCTION

Artificial general intelligence (AGI) refers to systems capable of understanding, learning, and applying knowledge across a wide range of tasks and domains at a level comparable to or exceeding that of humans [1]. Despite decades of research, AGI has not yet been realized. To advance toward AGI, we propose a novel perspective through an **Intelligent Foundation Model** (IFM). The main idea of IFM is to learn “intelligence” from a broad range of intelligent behaviors, rather than specific manifestations such as language, vision, or other cognitive tasks. To date, the only known instance of such general intelligence is human intelligence [2]. IFM therefore aims to capture human intelligence by learning from diverse human intelligent behaviors to approach AGI. This approach is grounded in the observation that every human intelligent behavior, whether involving perception, reasoning, or decision-making, can ultimately be represented as a temporal sequence of neuronal input–output transformations in human brain. By framing intelligence in terms of these observable sequences, IFM transforms the abstract and elusive goal of “learning intelligence” into a concrete sequence learning problem. Following

B. Cai is with Hangzhou International Innovation Institute, Beihang University, China. E-mail: caibr@buaa.edu.cn.

Y. Zhao is with RMIT University, Melbourne, Australia. E-mail: zhaoya514@gmail.com.

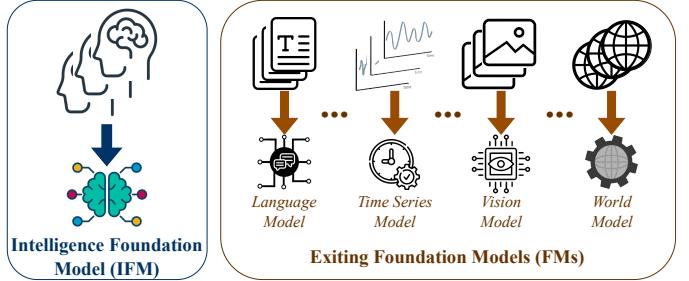


Fig. 1. Comparison between IFM and existing FMs.

the paradigm of FMs [3], IFM utilizes large-scale neuronal input–output transformation data that comprehensively represent diverse human intelligent behaviors. From these data, we expect that IFM can learn the universal underlying mechanisms that give rise to human cognitive abilities, enabling artificial systems to generalize across tasks and domains and move beyond narrow artificial intelligence toward true AGI.

**Fundamental barriers to AGI for existing FMs:** Existing FMs adhere to the paradigm of large-scale, task-agnostic pretraining to learn universal representations that can be efficiently adapted to diverse downstream tasks and modalities [4]. One of the examples is large language models (LLMs), e.g., ChatGPT [5], DeepSeek [6], and Gemini [7], which approximate human ability in text processing and have demonstrated impressive capabilities across a wide range of natural language tasks, e.g., question answering [8], translation [9], summarization [10], and code generation [11]. However, existing FMs constantly face limitations across various intelligent capabilities, including but not limited to:

- **Reasoning:** they often generate unreliable reasoning results, producing factually incorrect or fabricated content with apparent confidence [12].
- **Learning:** they cannot improve or update their behavior based on new experiences or interactions, and rely entirely on retraining or fine-tuning to adjust their responses [13].
- **Memory:** they cannot memorize new knowledge by themselves and must rely on external databases or retrieval systems to store and access updated information [14].

Overall, the core limitation of existing FMs is that they fail to represent the full spectrum of intelligence due to training to predict patterns in a single domain [15]. For example, LLMs fail at learning because language alone serves as a descriptive medium rather than a functional substrate of learning. While linguistic representations can articulate the principles of learning, they cannot embody the complex neural dynamics through which biological learning actually occurs.

As shown in Fig. 1, LLMs only learn statistical regularities and contextual dependencies in human language to predict and generate coherent text; time series FMs capture temporal dependencies in financial [16] and weather data [17]; vision FMs model spatial, temporal, and semantic continuity in images [18] and videos [19]; and world models [20] learn the dynamics and causal relationships of physical environments. In general, existing FMs primarily capture the partial expressions or surface regularities of intelligence, and thus fail to capture its underlying mechanisms comprehensively [21], [22].

**Key advantages of IFM for AGI:** Unlike existing FMs that are designed for specific modalities such as language or vision, IFM learns directly from diverse human behaviors that embody broad intelligence. Intelligence, however, has long been an elusive subject, and numerous theoretical frameworks have attempted to explain its nature. For example, predictive processing theories, including Free Energy Principle (FEP) [23] and the predictive mind hypothesis [24], conceptualize the brain as a predictive machine that seeks to minimize the discrepancy between internal predictions and sensory observations. Meanwhile, information-integration theories, such as Global Workspace Theory (GWT) [25], emphasize that intelligence emerges from the integration of specialized modules that share and broadcast information across a global neural workspace [26]. Other approaches focus on more localized mechanisms, such as the complementary learning system frameworks, which highlight the distinct yet cooperative roles of the hippocampus and neocortex in forming long- and short-term memories [27], or studies that track how the hippocampus and striatum encode value signals during decision-making [28]. While these theories offer valuable insights into specific aspects of cognition, most focus on explaining why intelligence arises rather than how to model and replicate it in machines at scale for building AGI.

It is widely acknowledged that human intelligence arises from the collective dynamics of approximately 86 billion neurons in the brain [29]. However, the immense complexity of these interactions makes it nearly impossible to construct an explicit, mechanistic model of intelligence from first principles. To address this challenge, IFM is designed to capture the implicit structural neuronal activity that underlies observed intelligent behaviors, learning directly from large-scale neuronal data reflecting diverse human cognitive processes. By grounding itself in observable behaviors rather than abstract theoretical formulations, IFM avoids reliance on hand-crafted models of intelligence and offers a more practical pathway toward AGI.

**Conceptual Overview of IFM:** As illustrated in Fig. 2, IFM learns the implicit structural neuronal activity that underlies observed intelligent behaviors. It achieves this through two core designs:

- *State Neural Network:* IFM is constructed upon a state neural network designed to emulate the dynamic behavior of biological neurons. This design encompasses three key aspects. First, each state neuron maintains an internal state that stores historical information for future decision-making

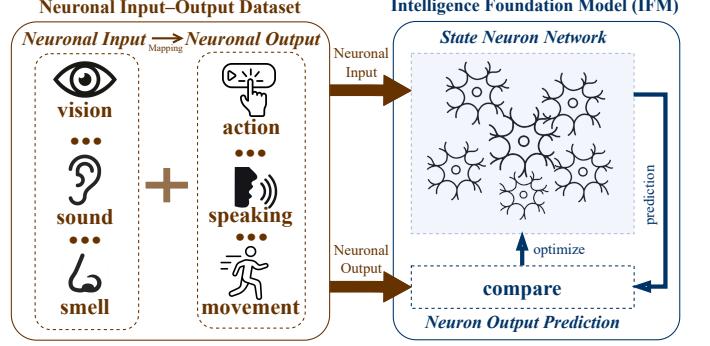


Fig. 2. Conceptual overview of IFM.

(*Neuron Function*). Second, the neurons are interconnected through recurrent, graph-like connections that allow information to branch, merge, and loop across multiple pathways, supporting a wide spectrum of intelligent functions (*Neuron Connectivity*). Third, connection strengths are allowed to evolve based on neuronal activity, enabling the network to self-organize, adapt, and continuously refine its internal structure through experience (*Neuron Plasticity*).

- *Neuron Output Prediction:* The collective dynamics of neurons underlying diverse human intelligent behaviors are formalized as the process of generating neuronal outputs from temporal neuronal input signals of the biological neuron system. Following this principle, IFM formulates its learning objective as predicting biological neuronal outputs from corresponding input dynamics over time, thereby enabling the model to internalize the essential computational processes that underpin intelligence. This objective directly links low-level neuronal activity to high-level intelligent behaviors, enabling the model to learn temporal, recurrent, and context-dependent transformations that characterize human intelligence. In essence, it bridges neural dynamics with the mechanisms of intelligence.

Equipped with these designs, IFM can be trained using structured datasets that encode intelligent behaviors as temporal sequences of neuronal inputs and outputs, allowing the model to learn through standard backpropagation techniques.

## II. IFM TECHNICAL DETAILS

IFM learns the implicit structural neuronal activity that underlies observed intelligent behaviors. We detail its design as follows.

### A. State Neural Network

We define state neural network as  $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ , where  $\mathcal{N} = \{\eta_1, \eta_2, \dots, \eta_n\}$  denotes the state neurons in  $\mathcal{G}$ , and  $\mathcal{E}$  denotes the edges that connect these state neurons. We categorize state neurons as **input neurons**, **output neurons**, and **hidden neurons** as follows.

- *Input neurons*  $\mathcal{N}_{in} \in \mathcal{N}$  are neurons that can receive stimulus signals originating from outside the system, analogous to human sensory neurons responsible for processing visual, auditory, olfactory, and other external inputs.

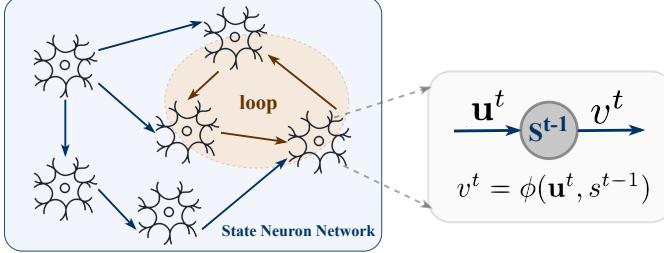


Fig. 3. State neural network.

- *Output neurons*  $\mathcal{N}_{out} \in \mathcal{N}$  are neurons whose activity can be externally perceived. This typically includes neurons that drive observable actions—such as motor neurons that control muscle contraction, and certain neurons whose activity reflects internal states of  $\mathcal{G}$ , providing insight into the network’s internal dynamics.
- *Hidden neurons*  $\mathcal{N}_h \in \mathcal{N}$  are neurons that receive signals only from within the network and whose activity does not directly produce observable outputs. They process and integrate these signals, maintain the network’s internal state, and mediate information flow between input and output neurons, enabling the system to perform complex computations.

With these neurons, state neural network  $\mathcal{G}$  receives an input stimulus sequence with input neurons  $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^T)$ ,  $\forall \mathbf{x}^t \in \mathbb{R}^{|\mathcal{N}_{in}|}$  over time, and generates an output response sequence through output neurons  $(\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^T)$ ,  $\forall \mathbf{y}^t \in \mathbb{R}^{|\mathcal{N}_{out}|}$ . By that, the function of  $\mathcal{G}$  can be defined as  $\mathbf{y}^t = \mathcal{G}(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^t)$ ,  $\forall 1 \leq t \leq T$ , as shown in Fig. 3.

The design of the state neural network  $\mathcal{G}$  models the neuron dynamics of biological neural systems, and it is built around three key components: **neuron function**, **neuron connectivity**, and **neuron plasticity**. Neuron function captures the internal state and temporal evolution of biological neurons, allowing each neuron to integrate signals over time and retain memory of past activations. Neuron connectivity models the flexible and recurrent wiring of the brain neuron cells, enabling signal flows through complex graphs to support oscillations, synchronization, and context-dependent computation. Neuron plasticity captures the adaptive adjustment of connection strengths in response to neuronal activity, reflecting biological learning mechanisms. Together, these three components establish a computational framework that embodies the dynamic, adaptive, and self-organizing nature of intelligence. They are detailed as follows.

**Neuron Function:** The function of a state neuron  $\eta \in \mathcal{N}$  is to maintain a dynamic internal state  $s$  that integrates past information and transforms incoming signals into outputs over time, thereby enabling temporal and context-dependent behavior. This process can be formalized as  $v^t = \phi(u^t, s^{t-1})$ , where  $u^t$  and  $v^t$  denote the input and output at time  $t$ , respectively. In this formulation, the neuron state  $s^t = \psi(u^t, s^{t-1})$  recursively integrates information from all previous time steps up to  $t$ . Recurrent [30] and spiking neurons [31] can be regarded as two special cases within this general structure. Specifically, a recurrent neuron produces continuous outputs, with

its internal state updated through a nonlinear transformation  $s^t = \sigma(W_{in}u^t + W_s s^{t-1})$ , where  $W_{in}$  and  $W_s$  are the input and recurrent weight matrices, respectively. In contrast, a spiking neuron generates discrete spikes, with its internal state governed by membrane potential dynamics. For example, in the Leaky Integrate-and-Fire (LIF) model [32], the membrane potential evolves as  $\frac{ds(t)}{dt} = -(s(t) - s_{rest}) + \mathbf{w}^\top \mathbf{u}(t)$ , where  $s_{rest}$  is the resting potential and  $\mathbf{w}$  denotes the synaptic weights for each input.

**Neuron Connectivity:** In the state neuron network  $\mathcal{G}$ , neurons are interconnected through directed, weighted edges represented by an adjacency matrix  $\mathcal{E} \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|}$ , where  $e_{ij}$  denotes the weight of the edge from neuron  $\eta_i$  to neuron  $\eta_j$ . Unlike the layer-wise connectivity typically adopted in existing FMs, our approach employs a graph-structured state neural network. This flexible connectivity supports both feedforward pathways and recurrent loops, allowing information to branch, merge, and circulate throughout the network. Consequently, signals can interact dynamically across time and multiple pathways, enabling the model to capture temporal dependencies, integrate heterogeneous information, and coordinate complex behaviors. These are capabilities that traditional layer-wise architectures inherently struggle to achieve.

Note that recurrent connectivity is essential for modeling human intelligence, yet it is largely overlooked in existing FMs. Fundamentally, recurrent connectivity allows information to persist and interact over time, creating temporal continuity that is critical for integrating past experiences, forming context-dependent representations, and coordinating complex behaviors. At the lowest level, two reciprocally connected neurons can form an oscillatory loop that generates periodic signals, similar to pacemaker neurons whose intrinsic rhythms coordinate neural processes throughout the brain [33]. When scaled up, larger neuronal assemblies sustain rhythmic firing patterns that synchronize activity across populations, enabling higher-order cognition [34]. These dynamics mirror the neural oscillatory activity, such as gamma and theta brain waves (generated by excitatory–inhibitory loops in the cortex and hippocampus) [35].

We believe that such recurrent structures will also play a crucial role in the emergence of machine consciousness. We conceptualize machine consciousness as the experience arising from continuous neural dynamics sustained by ongoing loops of information flows within the system. Through this recurrent process, external information gathered via sensory inputs is continuously integrated, giving rise to a coherent and dynamically evolving experience. Apparently, LLM does not have such machine consciousness.

**Neuron Plasticity:** Neuron plasticity is a fundamental mechanism in biological neurons, supporting essential capabilities such as learning and memory [36]. In biological neural systems, the strength and efficacy of synaptic connections can change over time in response to experience, sensory inputs, or environmental demands. This adaptability allows neural circuits to encode information, form memories, and reorganize themselves to optimize performance [37].

For two neurons in the state neural network  $\mathcal{G}$ ,  $\eta_p$  and  $\eta_q$ , that are connected by an edge of weight  $e_{pq}$ , we define  $\eta_p$  as the source (presynaptic) neuron that transmits a signal, and  $\eta_q$  as the target (postsynaptic) neuron that receives it. We see the core of neuron plasticity as the edge weights are affected by the historical activities of connected neurons. That is, the weight  $e_{pq}$  is influenced by the outputs of both neurons, i.e.,  $v_p$  of  $\eta_p$  (or  $u_q$  of  $\eta_q$ , equivalently) and  $v_q$  of  $\eta_q$ . Then, we have  $e_{pq}^t = \psi_e(u_q^1, \dots, u_q^t, v_q^1, \dots, v_q^t)$ . Since  $v_q = \phi(u_q, s_q)$ , we integrate it and the weight update becomes  $e_{pq}^t = \psi_e(u_q^t, e_{pq}^{t-1})$ , with  $e_{pq}^{t-1}$  as the previous edge weight. This can easily extend to target neurons with multiple inputs, i.e.,  $e_{pq}^t = \psi_e(\mathbf{u}_q^t, e_{pq}^{t-1})$ .

Many existing neural plasticity models fit such a formulation. For example, the widely adopted Spiking-Timing-Dependent-Plasticity (STDP) [38] in spiking neural networks [39] can be fitted as  $e_{pq}^t = \psi_e(\mathbf{u}_q^t, e_{pq}^{t-1}) = e_{pq}^{t-1} + \Delta e_{pq}^t$ , with  $\Delta e_{pq}^t$  determined by:

- *Long-term Potentiation (LTP):* If  $\eta_p$  fires before  $\eta_q$  (within a certain time window),  $\Delta e_{pq}^t$  is a positive weight increase.
- *Long-term Depression (LTD):* If  $\eta_p$  fires after  $\eta_q$ ,  $\Delta e_{pq}^t$  is a negative weight decrease.

**Summary:** With the above neuron function, neuron connectivity and neuron plasticity, the model of the entire state neural network becomes  $\mathbf{y}^t = \mathcal{G}(\mathbf{x}^t, (\mathbf{s}^{t-1}, \mathbf{e}^{t-1}))$ ,  $\forall 1 \leq t \leq T$ .  $\mathbf{s}^{t-1}$  represents states of all neurons and  $\mathbf{e}^{t-1}$  all edge weights at time  $t-1$ , with  $\mathbf{s}^t = \psi(\mathbf{u}^t, \mathbf{s}^{t-1})$  and  $\mathbf{e}^t = \psi_e(\mathbf{u}^t, \mathbf{e}^{t-1})$ .

## B. Neuron Output Prediction

Every manifestation of intelligent behavior, whether it involves pattern recognition, memory formation, causal reasoning, or decision-making, emerges from the transformation of neuronal inputs into outputs through complex and adaptive neural dynamics. Inspired by this principle, we define a neuron output prediction objective for IFM, enabling it to learn from intelligent behaviors represented as neuronal input–output sequences. This objective unifies diverse forms of intelligent behavior within a single learning framework, allowing IFM to capture the general mechanisms underlying intelligence.

**Formal Definition:** With ground truth input stimulus sequence  $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^T)$  and the response sequence  $(\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^T)$  sampled from real-world human intelligent behaviors, the formulation of the neuron output prediction is defined as  $\mathcal{L} = \min_{\mathcal{G}} \sum_{t=1}^T \ell(\mathbf{y}^t, \hat{\mathbf{y}}^t)$ , with  $\hat{\mathbf{y}}^t = \mathcal{G}(\mathbf{x}^t, (\mathbf{s}^{t-1}, \mathbf{e}^{t-1}))$ . In the objective,  $\mathcal{L}$  represents the total loss accumulated over a sequence of  $T$  time steps, and  $\ell(\mathbf{y}^t, \hat{\mathbf{y}}^t)$  measures the discrepancy between the real neuronal output  $\mathbf{y}^t$  and predicted output by IFM  $\hat{\mathbf{y}}^t$  at time  $t$ . By minimizing  $\mathcal{L}$ , IFM learns to predict correct neuronal outputs over time, capturing the temporal patterns and dependencies underlying intelligent behavior.

**Concept Illustration:** We illustrate the neuron output prediction concept using Pavlov's classical conditioning experiment [40], which is a foundational example of biological learning behavior [41]. We first sample neuronal input–output

sequences from the entire conditioning process, and then briefly explain how predicting neuronal outputs enables a state neural network to capture this biological-like learning behavior.

As illustrated in Fig. 4, the experiment shows that a dog learns to associate the sound of a ringing bell with the presentation of food through repeated training, eventually salivating at the sound alone. For simplicity, we model the dog with three functional neurons: a vision neuron that detects food ( $F$ ), a sound neuron that perceives ringing ( $R$ ), and a salivation neuron that triggers salivation ( $S$ ). The overall conditioning process has three stages:

- *Initial Stage:* The dog salivates when it sees food but not when it only hears the bell, i.e.,  $\{\mathbf{x}^1 = (F, \neg R), \mathbf{x}^2 = (\neg F, R)\}_{init}, \{\mathbf{y}^1 = S, \mathbf{y}^2 = \neg S\}_{init}$ .
- *Training Stage:* The dog simultaneously sees food and hears the bell over two trials, i.e.,  $\{\mathbf{x}^3 = (F, R), \mathbf{x}^4 = (F, R)\}_{train}, \{\mathbf{y}^3 = S, \mathbf{y}^4 = S\}_{train}$ .
- *Testing Stage:* The dog has learned to associate the bell with food, i.e.,  $\{\mathbf{x}^5 = (\neg F, R)\}_{test}, \{\mathbf{y}^5 = S\}_{test}$ .

Therefore, the entire conditioning process above can be represented as the temporally ordered sequences of neuronal inputs and outputs across the three stages combined, i.e.,  $\mathbf{x} = \{(F, \neg R), (\neg F, R), (F, R), (F, R), (\neg F, R)\}$  and  $\mathbf{y} = \{S, \neg S, S, S, S\}$ . More neuronal sequences can be obtained by varying the trial patterns, such as increasing or decreasing the number of trials in each stage, to enrich the dataset and improve sample diversity.

Under the objective of neuron output prediction, a state neural network is expected to be trained to generate proper  $\mathbf{y}$  given  $\mathbf{x}$ , indicating the possession of the conditioning-like learning ability. Specifically, we train the state neural network  $\mathcal{G}$  using the above neuronal input–output sequences, where two input neurons receive  $\mathbf{x}$  and one output neuron generates  $\mathbf{y}$ . The network includes multiple hidden neurons to learn the mapping from  $\mathbf{x}$  to  $\mathbf{y}$  by minimizing the neuron output prediction loss  $\mathcal{L}$ . Through this training process,  $\mathcal{G}$  adjusts its internal parameters that regulate neuron function, neuron connectivity, and neuron plasticity to fit the conditioning mechanism and reproduces similar neuronal output patterns during inference. Note that the trained network is not intended to memorize specific training samples, but rather to learn the fundamental principles of classical conditioning that underlie a large number of training samples.

## III. IFM TRAINING

To train IFM under the neuron output prediction objective, a major challenge lies in constructing neuronal input–output datasets that capture the temporal evolution of neural activity underlying intelligent behaviors. Since IFM is designed to learn the time-dependent mapping between neuronal inputs and outputs, its training data should ideally represent the dynamic transformation of signals as the neural system interacts with its environment. Each data sample consists of a pair of neuronal input–output sequences, expressed as  $d = \{(\mathbf{x}^1, \dots, \mathbf{x}^T), (\mathbf{y}^1, \dots, \mathbf{y}^T)\}$ , while the overall dataset comprises multiple such samples, denoted by

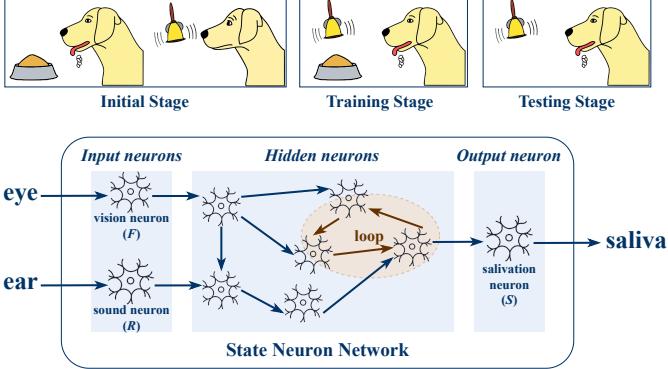


Fig. 4. Pavlov’s classical conditioning experiment.

$D = \{d_1, d_2, \dots, d_n\}$ . Following the paradigm of FMs, data samples should be collected from a wide range of individuals engaged in diverse intelligent behaviors. Such data samples can be obtained in two ways:

- *Direct Neuronal Sampling*: This approach acquires neuronal input–output data directly from biological brains by recording synaptic activity across neurons as subjects perform intelligent behaviors. It provides the most biologically faithful supervision for IFM, as it reflects the genuine neuronal computations underlying cognition. However, large-scale data collection remains technically infeasible because capturing the full dynamics of even small cortical circuits at cellular resolution remains an open challenge. Current techniques, such as Neuropixels probes [42] and calcium imaging [43], can only record limited subsets of neurons and modalities, making this method suitable primarily for small-scale or proof-of-concept IFMs.
- *Indirect Neuronal Sampling*: This scalable alternative captures functional equivalents of neuronal inputs and outputs through fully instrumented, human-in-the-loop embodiments, such as a neurosensory interface (a wearable multimodal sensing system) or an avatar (a digital or robotic agent with human-like sensing and actuation). These embodiments perceive environmental stimuli (visual, auditory, tactile, proprioceptive) and record observable behavioral responses (speech, gestures, motor actions). The sensory streams serve as proxies for neuronal inputs, while behavioral responses represent neuronal outputs. Recording paired temporal sequences during natural interactions produces high-dimensional data encoding the dynamics of intelligent behavior. Training IFM on such data enables it to learn how sensory experiences evolve into actions over time, effectively capturing neuron-like transformation principles without requiring direct observation of biological neurons.

As neural recording technologies progress, direct and indirect neuronal sampling are expected to converge, enabling IFM to move beyond emulating intelligent behavior toward internalizing its underlying mechanisms, thereby closing the loop between cognition and computation. With these data samples, IFM implemented by differentiable state neural networks can be efficiently trained via backpropagation. Given the streaming nature of the training data, variants of Trun-

cated Backpropagation Through Time (TBPTT) [44] can be employed to reduce computational complexity. We present a *toy example* in which IFM is used to train an artificial Pong player, demonstrating the overall design process available at [https://github.com/brcai/IFM\\_pong](https://github.com/brcai/IFM_pong).

#### IV. DISCUSSION AND OUTLOOK

**Discussion:** Although humans differ across age, culture, and individual biology, intelligence consistently arises from shared fundamental principles embedded in the brain. Building on this premise, IFM leverages the foundation model paradigm to capture these underlying mechanisms directly from neuronal dynamics. This approach defines a fundamentally new pathway toward AGI, one that models the mechanisms of intelligence itself rather than its fragmented expressions. We envisage a pragmatic and scalable roadmap toward AGI through IFM from two complementary perspectives:

- *The first trajectory* focuses on biological scaling, starting from simpler neural systems or lower organisms and progressively advancing to more complex brains, to systematically learn the principles of intelligence across increasing levels of complexity;
- *The second trajectory* focuses on functional scaling, developing IFMs for specific robotic applications such as industrial manipulators, household assistants or autonomous agents and gradually integrating these specialized capabilities into a unified, general-purpose intelligent system.

By integrating these two trajectories, IFM offers a stepwise, interpretable, and biologically inspired pathway toward AGI.

**Outlook:** A key implication of IFM is its potential in conceptual unification of human intelligence and machine intelligence. By modeling the universal principles that generate intelligence, IFM’s artificial substrates (state neural networks) can replicate the functional dynamics of human cognition without demanding any specific biological neuron cells. The success of IFM will show that consciousness, learning, reasoning, and memory emerge from dynamic, recurrent, and plastic interactions, whether instantiated in biological or artificial media. This opens the door to transformative possibilities, such as continuous mind uploading [45], where biological neurons are gradually replaced or augmented with artificial state neurons while preserving the continuity of consciousness. Such an approach safeguards personal identity and cognitive continuity, surpassing traditional “copy-based” mind uploading paradigms [46], and reframes our understanding of intelligence as substrate-independent. That paints a transformative vision for the future of intelligent life, where the boundary between natural and artificial intelligence becomes increasingly seamless and where human-level cognition can be instantiated beyond the biological brain.

IFM envisions a future where intelligence is no longer confined to biology. Humans and machines could share a common cognitive substrate, enabling seamless augmentation, collaboration, and extension of minds. Adaptive artificial neurons could integrate with biological networks, expanding memory, perception, and reasoning capacities beyond natural limits. In this new paradigm, consciousness, learning, and creativity

are universal phenomena that can be instantiated in multiple substrates, offering humanity unprecedented opportunities to explore, understand, and even evolve intelligence itself. IFM does not merely aim to replicate intelligence, but it lays the foundation for a future in which the boundaries between human, machine, and mind are dynamically intertwined, ushering in a new era of cognitive life.

## REFERENCES

- [1] N. Fei, Z. Lu, Y. Gao, G. Yang, Y. Huo, J. Wen, H. Lu, R. Song, X. Gao, T. Xiang *et al.*, “Towards artificial general intelligence via a multimodal foundation model,” *Nature Communications*, vol. 13, no. 1, p. 3094, 2022.
- [2] S. Russell, P. Norvig, and A. Intelligence, “A modern approach,” *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, vol. 25, no. 27, pp. 79–80, 1995.
- [3] M. Awais, M. Naseer, S. Khan, R. M. Anwer, H. Cholakkal, M. Shah, M.-H. Yang, and F. S. Khan, “Foundation models defining a new era in vision: a survey and outlook,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.
- [4] R. Bommasani, “On the opportunities and risks of foundation models,” *arXiv preprint arXiv:2108.07258*, 2021.
- [5] OpenAI, “Chatgpt,” <https://chat.openai.com>, 2023.
- [6] A. Liu, B. Feng, B. Xue, B. Wang, B. Wu, C. Lu, C. Zhao, C. Deng, C. Zhang, C. Ruan *et al.*, “Deepseek-v3 technical report,” *arXiv preprint arXiv:2412.19437*, 2024.
- [7] G. Team, P. Georgiev, V. I. Lei, R. Burnell, L. Bai, A. Gulati, G. Tanzer, D. Vincent, Z. Pan, S. Wang *et al.*, “Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context,” *arXiv preprint arXiv:2403.05530*, 2024.
- [8] M. Yue, “A survey of large language model agents for question answering,” *arXiv preprint arXiv:2503.19213*, 2025.
- [9] Z. He, T. Liang, W. Jiao, Z. Zhang, Y. Yang, R. Wang, Z. Tu, S. Shi, and X. Wang, “Exploring human-like translation strategy with large language models,” *Transactions of the Association for Computational Linguistics*, vol. 12, pp. 229–246, 2024.
- [10] H. Zhang, P. S. Yu, and J. Zhang, “A systematic survey of text summarization: From statistical methods to large language models,” *ACM Computing Surveys*, vol. 57, no. 11, pp. 1–41, 2025.
- [11] J. Jiang, F. Wang, J. Shen, S. Kim, and S. Kim, “A survey on large language models for code generation,” *arXiv preprint arXiv:2406.00515*, 2024.
- [12] M. Mahaut, L. Aina, P. Czarnowska, M. Hardalov, T. Müller, and L. Márquez, “Factual confidence of llms: on reliability and robustness of current estimators,” *arXiv preprint arXiv:2406.13415*, 2024.
- [13] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen *et al.*, “Lora: Low-rank adaptation of large language models,” *ICLR*, vol. 1, no. 2, p. 3, 2022.
- [14] W. Zhong, L. Guo, Q. Gao, H. Ye, and Y. Wang, “Memorybank: Enhancing large language models with long-term memory,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 17, 2024, pp. 19 724–19 731.
- [15] R. T. McCoy, S. Yao, D. Friedman, M. Hardy, and T. L. Griffiths, “Embers of autoregression: Understanding large language models through the problem they are trained to solve,” *arXiv preprint arXiv:2309.13638*, 2023.
- [16] B. A. Marconi, “Time series foundation models for multivariate financial time series forecasting,” *arXiv preprint arXiv:2507.07296*, 2025.
- [17] S. Chen, G. Long, J. Jiang, D. Liu, and C. Zhang, “Foundation models for weather and climate data understanding: A comprehensive survey,” *arXiv preprint arXiv:2312.03014*, 2023.
- [18] Z. Chen, J. Wu, W. Wang, W. Su, G. Chen, S. Xing, M. Zhong, Q. Zhang, X. Zhu, L. Lu *et al.*, “Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 24 185–24 198.
- [19] K. Li, Y. Wang, Y. Li, Y. Wang, Y. He, L. Wang, and Y. Qiao, “Unmasked teacher: Towards training-efficient video foundation models,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 19 948–19 960.
- [20] A. Bar, G. Zhou, D. Tran, T. Darrell, and Y. LeCun, “Navigation world models,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 15 791–15 801.
- [21] E. Fedorenko, S. T. Piantadosi, and E. A. Gibson, “Language is primarily a tool for communication rather than thought,” *Nature*, vol. 630, no. 8017, pp. 575–586, 2024.
- [22] S. Pinker, *The language instinct: How the mind creates language*. Penguin uK, 2003.
- [23] K. Friston, “The free-energy principle: a unified brain theory?” *Nature reviews neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.
- [24] A. Clark, “Whatever next? predictive brains, situated agents, and the future of cognitive science,” *Behavioral and brain sciences*, vol. 36, no. 3, pp. 181–204, 2013.
- [25] B. J. Baars, *A cognitive theory of consciousness*. Cambridge University Press, 1993.
- [26] M. Shanahan, *Embodiment and the inner life: Cognition and Consciousness in the Space of Possible Minds*. Oxford University Press, 2010.
- [27] R. C. O'Reilly and K. A. Norman, “Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework,” *Trends in cognitive sciences*, vol. 6, no. 12, pp. 505–510, 2002.
- [28] G. E. Wimmer and D. Shohamy, “Preference by association: How memory mechanisms in the hippocampus bias decisions,” *Science*, vol. 338, no. 6104, pp. 270–273, 2012.
- [29] S. Herculano-Houzel, “The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost,” *Proceedings of the National Academy of Sciences*, vol. 109, no. supplement\_1, pp. 10 661–10 668, 2012.
- [30] Y. Yu, X. Si, C. Hu, and J. Zhang, “A review of recurrent neural networks: Lstm cells and network architectures,” *Neural computation*, vol. 31, no. 7, pp. 1235–1270, 2019.
- [31] S. Ghosh-Dastidar and H. Adeli, “Spiking neural networks,” *International journal of neural systems*, vol. 19, no. 04, pp. 295–308, 2009.
- [32] X. Yao, F. Li, Z. Mo, and J. Cheng, “Glif: A unified gated leaky integrate-and-fire neuron for spiking neural networks,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 32 160–32 171, 2022.
- [33] J. A. Mohawk, C. B. Green, and J. S. Takahashi, “Central and peripheral circadian clocks in mammals,” *Annual review of neuroscience*, vol. 35, no. 1, pp. 445–462, 2012.
- [34] P. Fries, “Rhythms for cognition: communication through coherence,” *Neuron*, vol. 88, no. 1, pp. 220–235, 2015.
- [35] G. Buzsaki and A. Draguhn, “Neuronal oscillations in cortical networks,” *science*, vol. 304, no. 5679, pp. 1926–1929, 2004.
- [36] T. V. Bliss and T. Lømo, “Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path,” *The Journal of physiology*, vol. 232, no. 2, pp. 331–356, 1973.
- [37] R. C. Malenka and M. F. Bear, “Ltp and ltd: an embarrassment of riches,” *Neuron*, vol. 44, no. 1, pp. 5–21, 2004.
- [38] N. Caporale and Y. Dan, “Spike timing-dependent plasticity: a hebbian learning rule,” *Annu. Rev. Neurosci.*, vol. 31, no. 1, pp. 25–46, 2000.
- [39] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.
- [40] P. I. Pavlov, “Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex,” *Annals of neurosciences*, vol. 17, no. 3, p. 136, 2010.
- [41] J. E. Mazur, *Learning and behavior: Instructor's review copy*. Psychology Press, 2015.
- [42] J. J. Jun, N. A. Steinmetz, J. H. Siegle, D. J. Denman, M. Bauza, B. Barbarits, A. K. Lee, C. A. Anastassiou, A. Andrei, Ç. Aydin *et al.*, “Fully integrated silicon probes for high-density recording of neural activity,” *Nature*, vol. 551, no. 7679, pp. 232–236, 2017.
- [43] T.-W. Chen, T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. A. Kerr, M. B. Orger, V. Jayaraman *et al.*, “Ultrasensitive fluorescent proteins for imaging neuronal activity,” *Nature*, vol. 499, no. 7458, pp. 295–300, 2013.
- [44] R. J. Williams and J. Peng, “An efficient gradient-based algorithm for on-line training of recurrent network trajectories,” *Neural computation*, vol. 2, no. 4, pp. 490–501, 1990.
- [45] R. W. Clowes and K. Gärtner, “Slow continuous mind uploading,” in *The Mind-Technology Problem: Investigating Minds, Selves and 21st Century Artefacts*. Springer, 2021, pp. 161–183.
- [46] G. Piccinini, “The myth of mind uploading,” in *The mind-technology problem: Investigating minds, selves and 21st century artefacts*. Springer, 2021, pp. 125–144.