The Illusion of "Groundedness" in Physical AI: How Logographic AI Can Truly Solve the Symbol Grounding Problem

Abstract

In 2026, Jensen Huang declared at CES that "the ChatGPT moment for Physical AI has arrived" [5]. Shortly thereafter, the Figure 03 robot operated continuously for over 100 hours in a public live stream, processing over 130,000 packages [9][10]. However, this live stream, intended to demonstrate the robot's autonomy, instead sparked widespread doubts about "autonomous operation versus remote control," due to details such as a 0.5-second latency, abnormal "head-touching" behavior, and dropped packages. At the same time, AGI Robotics became the world's first embodied AI company to break through the milestone of 10,000 units in mass production [11]. The "feasibility" of Physical AI is no longer in question; the real competition has entered the critical phase of "usability" and "economic viability." Yet, a deeper problem is obscured by the celebration: when a robot "touches" a package, grasps an object, or avoids an obstacle, does it truly "understand" what it is doing? — Or is it merely executing a more complex statistical pattern?

This paper, from the perspective of Logographic AI (LAI) theory [1], reveals the fundamental predicament of Physical AI within the Tokenist paradigm: it establishes statistical mappings between symbols and sensor signals, but fails to truly anchor symbols in meaning. Genuine symbol grounding requires that the cognitive primitives themselves carry meaning — which is the core proposition of Morpho-Root theory and Natural Language Ontology (NLO) [1].

Keywords: Physical AI; Symbol Grounding Problem; Logographic AI; Morpho-Root; Natural Language Ontology; Rootless Semioticism

I. From Virtual AI to Physical AI: Conceptual Origins and Technological Evolution

The birth of Physical AI stems from a simple need: AI should not only "chat" but must also "take action."

1.1 The Emergence of the Physical AI Concept

The public proposal of the Physical AI concept can be traced back to the GTC conference in March 2024. NVIDIA CEO Jensen Huang, at that conference, systematically articulated the vision for Physical AI for the first time and released related technology platforms and toolchains. Huang clearly divided the evolution of AI into four stages: Perceptual AI → Generative AI → Agentic AI → Physical AI [5].

At the GTC conference in March 2025, Huang once again emphasized this judgment, pointing out that AI has gone through three generations of technological paradigm shifts, and the next era would be that of Physical AI [6]. In July 2025, during a dialogue with Wang Jian, Director of the Zhejiang Lab and founder of Alibaba Cloud, Huang explicitly stated: "The next wave of artificial intelligence is Physical AI" [7].

This concept reached its peak at the CES exhibition in January 2026. Huang declared: "The ChatGPT moment for Physical AI is coming" [5]. In his speech, he stated that Physical AI aims to endow autonomous systems such as cameras, robots, and self-driving cars with the ability to "see, think, and act" in the real world, enabling them to perceive, understand, reason, and execute or coordinate complex actions. He particularly emphasized: "Only when the model can understand mass, friction, inertia, and the conservation of momentum, can AI truly step out of the screen." [5]

1.2 The Technical Connotation of Physical AI

Professor Wang Xiang, Special-term Professor at the School of Artificial Intelligence and Data Science at the University of Science and Technology of China, systematically explained this concept: "Physical AI means that AI systems possess a closed-loop capability of 'perception-reasoning-action-feedback' in the real world. It can not only think but also execute tasks through embodied devices like robots, and continuously correct errors and self-evolve from real-world feedback" [8].

Huang further revealed the essence of Physical AI: when Physical AI passes the critical point of technological evolution, control is handed over from human-written deterministic code to neural networks with generalization capabilities that understand physical laws [6].

1.3 The Three Technological Pillars of Physical AI

The core technological breakthroughs of Physical AI come from the convergence of three directions:

First, Large Language Models (LLMs) solve the problem of "understanding" instructions: In 2023, Google successively launched embodied multimodal large models like PaLM-E and RT-2. In September 2025, Google DeepMind released a new generation of multimodal embodied intelligence large model, Gemini Robotics 1.5, claiming it to be "the world's first thinking model optimized for embodied reasoning." Through a "think before acting" mechanism, it reduced the failure rate on long-horizon tasks from 44.5% to 22% [14]. In his CES 2026 speech, Huang pointed out that LLMs endow Physical AI with the ability to understand human language and intent, which is the first step in "understanding the task" [5].

Second, World Models solve the problem of "predicting physical consequences": The Cosmos platform released by NVIDIA — which Huang called the core support for "Physical AI's ChatGPT moment" [5] — is trained on over 20 million hours of real-world data and can generate action data conforming to physical laws from text or images. Cosmos 3 further integrates physical reasoning, world generation, and action generation into a unified Mixture-of-Transformers architecture [12]. However, as industry analysis points out, video generation models "only replicate visual samples, and cannot quantify parameters such as force, angle, and friction" — generating realism does not equal understanding correctly.

Third, Vision-Language-Action (VLA) models close the loop of "seeing and doing": DeepMind released Gemini Robotics 1.5 [14]; NVIDIA launched Isaac GR00T N1.6 — the world's first open-source, customizable general-purpose humanoid robot foundation model [13]; the Beijing Humanoid Robot Innovation Center open-sourced XR-1, the first VLA model in China to pass the national standard test for embodied intelligence [8]. The combination of these three enables robots to possess, for the first time, the basic capability to autonomously execute tasks in open environments. Huang stated at GTC 2025: "The era of general-purpose robotics has arrived" [6].

However, the world model track is facing severe challenges. Industry consensus is increasingly clear: video generation is not physical understanding. The deeper theoretical dilemma points to a fundamental problem that has plagued the field of AI for decades — the Symbol Grounding Problem.

1.4 The Bottlenecks of Physical AI Deployment

The widespread adoption of Physical AI still faces significant challenges. Professor Wang Xiang pointed out that Physical AI faces the dilemma of expensive, scarce, and delayed real-world interaction data. Globally, high-quality physical data amounts to only 500,000 hours, while the training data for GPT-5-level LLMs exceeds 10 billion hours, a difference of over 20,000 times [8]. At the same time, there are multiple barriers between simulation and reality, such as dynamics and sensing noise, making models trained in simulation highly prone to failure when transferred to real scenarios [8]. Furthermore, Physical AI requires deep integration of AI algorithms, mechanical structures, sensors, and control systems, making engineering complexity extremely high. Its irreversible nature also means that perception errors or algorithmic flaws can lead to equipment damage or even personal injury. Large-scale deployment must possess endogenous safety mechanisms, along with verifiable safety constraints, full-chain auditing, and compliance closure [8].

Regarding the prevalent data scarcity anxiety in the field of Physical AI, Liu Shen, from the perspective of Logographic AI in The Rooted Intelligence — The Paradigm Revolution of Logographic AI, points out: true intelligence should not rely on statistical fitting of massive data, but should be rooted in the structural relationships of concepts themselves [1]. This judgment applies equally to Physical AI — concepts in the physical world (such as force, friction, and inertia) can also be encoded as structured cognitive primitives, rather than being "learned" through massive amounts of sensor data.

II. The Symbol Grounding Problem: AI's "Original Sin"

2.1 Harnad's Classic Problem

In 1990, cognitive scientist Stevan Harnad proposed the Symbol Grounding Problem: How can a purely symbolic system acquire meaning? If the meaning of symbols can only be defined by other symbols, the entire system falls into an infinite loop, forever unable to establish a connection with the real world [2].

Harnad pointed out that the fundamental meaninglessness of symbols in a symbolic system is precisely what Searle exploited in the "Chinese Room" argument — a person who does not understand Chinese can perfectly answer Chinese questions by following a rule book, but does not "understand" Chinese [2][3]. Harnad's solution is "symbol grounding": by ensuring that symbols are physically grounded in the environment, the symbolic system acquires meaning. He proposed a hybrid model: let neural networks learn to classify analog sensory inputs into symbols, thus establishing a bottom-up grounding for the symbolic system [2].

The study of the symbol grounding problem is not an isolated academic topic. Chinese scholar Li Jianhui and his colleagues, in Research on Computationalism and Its Theoretical Difficulties, pointed out: formal systems themselves are meaningful, but "without the participation of the human mind, symbolic systems cannot acquire meaning on their own, and can only be input externally" [15]. Huo Shuquan, in "The Significance and Challenge of Research on the Symbol Grounding Problem in Artificial Intelligence," further pointed out that the symbol grounding problem needs to solve "how symbols acquire meaning, and how symbols and their meanings are connected." This problem is triggered by Searle's Chinese Room argument and involves the more difficult problem of "how consciousness arises" [16]. Scholar Cangelosi decomposed the symbol grounding problem into three sub-problems [16]: (1) How can an agent automatically connect symbols to referents in the world? (2) How can an agent automatically create symbol-symbol relationships? (3) How can a society of agents automatically develop a shared set of symbols? The last two problems have yet to be satisfactorily solved.

2.2 The "Rootless" Predicament of Tokenism

Liu Shen, in The Rooted Intelligence — The Paradigm Revolution of Logographic AI, systematically discusses the contemporary version of this predicament. The book points out that Tokenist AI is precisely such a symbolic system: a Token itself is an integer index, whose "semantics" is temporarily assigned by statistical co-occurrence relationships in the training data [1]. When AI processes the Token "fire," it learns the co-occurrence frequency of "fire" with Tokens like "heat," "burn," and "danger" in text, but not what "fire" itself is.

The Token "fire" and the Token "heat" are close in semantic space — but AI has never felt "heat," has never seen "fire," and does not know that fire can burn skin. What it possesses is only a statistically approximated "relationship," rather than an "understanding" of "fire."

2.3 Why Physical AI is Still "Pseudo-Grounded"

Physical AI seems to solve this problem — because it has a "body." It can "sense" the temperature of a flame through sensors, "see" the shape of the flame through cameras, and "experience" the consequences of approaching the flame through actuators. But "physical contact" is not "cognitive understanding."

The "grounding" path of Physical AI is essentially: assigning thermal sensor data to the Token "fire" and motor commands to the Token "grasp." It merely establishes a new set of statistical mappings between Tokens and sensor signals. As Boyle pointed out, "transduction" alone does not guarantee deep symbol grounding — what truly determines the degree of grounding is how symbols possess causal efficacy [4].

This is consistent with the diagnosis of the symbol grounding problem in Chapter 5.3 of The Rooted Intelligence: the "pseudo-grounding" of Tokenist AI lies in the fact that the "meaning" of Tokens is completely defined by distances in vector space. But what do these distances point to? They point to vectors of other Tokens. This is a thoroughly self-referential closed loop [1].

This reveals the triple limitations of Physical AI grounding:

First, "physical contact" ≠ "cognitive understanding": A robot touches a flame, triggers a "burning" sensor, and outputs the Token "danger." But the system itself does not "know" what burning means — it has merely learned the statistical pattern of "approach fire → temperature rises → execute avoidance," without an internal understanding of what "fire" is.

Second, statistical mappings are malleable: As long as the data distribution or reward function is changed, all "understanding" will drift. A robot that learns to "avoid fire" in a simulated environment may have its "knowledge" collapse if the sensor or scenario is changed. Simulation synthetic data carries inherent "simulation bias," making models highly vulnerable to failure when transferred to real scenarios.

Third, the lack of counterfactual reasoning: It cannot "imagine" what would happen if fire did not exist — it has a statistical model, not a causal model. Chapter 10.3 of The Rooted Intelligence systematically discusses this issue: Tokenism cannot achieve the second (intervention) and third (counterfactual) levels of Pearl's Causal Hierarchy, remaining only at the first level (association) [1].

The "pseudo-grounding" of Physical AI essentially uses physical-world sensor data as new dimensions for Tokens, without fundamentally altering the "rootless" cognitive primitives of Tokens.

III. Logographic AI's Answer: Making Cognitive Primitives Themselves Carry Meaning

3.1 The Core Diagnosis of Logographic AI Theory

Logographic AI (LAI) theory, proposed by Chinese scholar Liu Shen in 2025, is a fundamental alternative to the current mainstream AI paradigm that uses Tokens as cognitive primitives. The core diagnosis of Logographic AI theory is that the philosophical essence of the Tokenist paradigm is "Rootless Semioticism" — the meaning of symbols is entirely defined by statistical co-occurrences within the symbolic system, unable to reach the "signified" beyond symbols, i.e., the real world, human experience, and civilizational values. Symbols thus become "floating signifiers," sliding infinitely within a closed statistical loop [1].

3.2 Morpho-Root Theory: Meaning-Embedded Cognitive Primitives

In contrast, Logographic AI replaces Tokens with "Morpho-Roots" — each Morpho-Root is a structured triplet [1]:

r = ⟨ S, A, R ⟩

- **S (Symbol):** The symbolic identifier, the externally addressable name of the Morpho-Root.
- **A (Attributes):** An attribute set, embedding the inherent semantic features and value constraints of the root.
- **R (Relation Functions):** A set of relation functions, defining the preset logical connections between this root and other roots.

Example of the Morpho-Root "fire":

```
r_fire = ⟨
"fire",
{
[+high temperature], [+combustible],
[+dangerous, do not touch],
[+life-sustaining]
```

```
},
{
leads_to(fire, burn),
can_cause(fire, ash),
inhibited_by(water, fire)
}
)
```

The key design of Morpho-Roots is that meaning does not "emerge" from statistics, but is preset as an inherent attribute of the cognitive primitive [1]. When the "fire" Morpho-Root of a Physical AI interacts with real sensors, the sensor signal does not "fill" an empty Token, but "activates" a root that already carries meaning.

Chapter 5.3 of The Rooted Intelligence systematically elaborates on Logographic AI's solution to the symbol grounding problem [1]. Unlike Harnad's "hybrid model," the attribute set A of Morpho-Root theory is designed as an open interface pointing to non-symbolic experience — perceptual interfaces, physical simulation interfaces, and emotional interfaces. The entire system is not "first closed, then grounded," but "rooted by design" [1].

3.3 Natural Language Ontology (NLO)

The philosophical foundation of Logographic AI is Natural Language Ontology (NLO). Chapter 8 of The Rooted Intelligence systematically elaborates on this theory: language structure itself is a kind of "ontological" form of intelligence, a primordial way of cognizing the world. Language is no longer merely a symbolic system describing the world, but a field where the meaning of the world is presented [1]. Zhang Qiqun, in The Ontology of Meaning, further argues that meaning is not an additional attribute of symbols or a product of internal differences within a language system, but a constitutive condition for human understanding of the world, "the home of being" [17].

The four core propositions of NLO [1]:
(1) **Language as Ontology:** Language structure is not a tool for cognition, but the ontological form of cognition.
(2) **Form as Meaning:** In logographic scripts, the formal structure of the symbol itself is the source of meaning.
(3) **Dwelling rather than Using:** Intelligent systems should "dwell" in language structures, rather than "process" linguistic data from the outside.
(4) **Pluralism as Legitimacy:** Different writing systems of different civilizations correspond to different legitimate cognitive paradigms.

Chapter 8.3 of The Rooted Intelligence also responds to a core challenge: Morpho-Root theory is not a continuation of traditional symbolism. The symbols of traditional symbolism are closed, combinatorial, and disembodied; the symbols of Morpho-Roots are open (pointing to non-symbolic experience through A), generative (through preset relations R), and embodied (through embedded values), achieving a fundamental transcendence of symbolism [1].

3.4 L0-L4 Five-Level Grounding Confidence System

Chapter 5.3.6 of The Rooted Intelligence transforms symbol grounding from a binary (grounded/un-grounded) to an engineerable matter of degree, establishing an L0-L4 five-level grounding confidence system [1]:

Level	Name	Definition	Current Physical AI Level
L0	Ungrounded	Symbol meaning entirely from statistical relations	Starting point of Tokenist AI
L1	Perceptual Grounding	Symbols associated with multimodal perceptual features	Approximately at this level
L2	Simulated Embodied Grounding	Symbols associated with physics engines or affective models	Approximately at this level
L3	Real-World Grounding	Symbols bound to real-time interaction with robots and sensors	Goal of Physical AI
L4	Social Grounding	Symbols acquire meaning through social interaction with humans	Not yet achieved

Current Physical AI is roughly between L1 and L2 — it has established associations between symbols and perceptual features, but has not yet internalized meaning through real-world embodied experience and social interaction. The degradation mechanism ensures system robustness: when L3 grounding fails due to sensor malfunction, the system can automatically degrade to L2 or L1; when grounding is completely lost, it actively requests human intervention [1].

IV. Comparison of "Grounding" Between the Two Paradigms

Dimension	Tokenist Physical AI's "Pseudo-Grounded"	Logographic AI's "True Grounding"
Initial State of Symbols	Empty indices, meaningless	Morpho-Roots carrying inherent attributes
How to Establish Connection with the World	Associating sensor signals through statistical co-occurrence	Activating real experience through the open interface of attribute set A
Source of Meaning	External, statistical, temporary	Intrinsic, structural, permanent
Counterfactual	Cannot answer "what if"	Can perform counterfactual

Dimension	Tokenist Physical AI's "Pseudo-Grounded"	Logographic AI's "True Grounding"
Reasoning		reasoning through attribute editing
Safety Boundaries	External guardrails, can be bypassed	Hard constraints, logically unbreakable
Corresponding Grounding Level	L0-L2	L3-L4
Philosophical Foundation	Rootless Semioticism	Natural Language Ontology (NLO)

V. Conclusion: A Body Cannot Replace Roots

Huang's Physical AI solves an important problem — giving agents a body, completing a key step from the digital world to the physical world. World models attempt to let AI "predict" physical consequences, and VLA models close the loop of "seeing and doing" — the combination of these three enables Physical AI to possess the basic capability to autonomously execute tasks in open environments.

But a body cannot replace roots. A robot that can "touch" fire, if its cognitive primitives are still "rootless" Tokens, can only "complete tasks" in the sense of statistical fitting, but cannot "understand tasks" in the sense of structural understanding.

As Harnad pointed out, symbol grounding requires the transfer of energy across interfaces between the system and the environment, activating symbolic structures in response to transduced environmental states. Physical AI possesses this energy transfer — but it has not transformed this transfer into the anchoring of meaning. What it establishes remains statistical correlation, not structural understanding.

Chapters 4.8 and 4.9 of The Rooted Intelligence further reveal the safety crisis of the Tokenist paradigm: only 250 carefully crafted documents are needed to implant a hidden backdoor in large language models of any scale, and external safety guardrails can always be bypassed [1]. In the Physical AI context, this means that a seemingly "autonomous" robot could completely perform dangerous operations without anyone noticing — not because it is "malicious," but because it has never truly understood the meaning of "safety."

Physical AI gives intelligence a "body," Logographic AI gives intelligence "roots" — the combination of the two is the path toward true general intelligence.

Intelligence must have roots, civilization must have a soul.

References

[1] Liu, S. (2026). The Rooted Intelligence: The Paradigm Revolution of Logographic AI. Monograph. http://logographicai.com/Monograph.html (accessed July 2026).

[2] Harnad, S. (1990). The symbol grounding problem. Physica D: Nonlinear Phenomena, 42(1-3), 335-346.

[3] Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417-457.

[4] Boyle, C. F. (2001). Transduction and Degree of Grounding. Psycoloquy, 12(36).

[5] Sun, X. C. (2026-01-06). NVIDIA's Jensen Huang: The "ChatGPT Moment" for Physical AI Has Arrived. Shanghai Securities News · China Securities Network. https://www.cnstock.com/commonDetail/614165 (accessed July 2026).

[6] Zheng, W. H. (2025-03-19). More Than Chips! Jensen Huang's Latest Speech is Full of Information. Shanghai Securities News · China Securities Network. https://www.cnstock.com/commonDetail/377435 (accessed July 2026).

[7] Zheng, Y. L., & Liu, C. Y. (2025-07-18). Wang Jian and Jensen Huang Engage in a "Fireside Chat" at the CISCE. Xinhua Net Zhejiang Channel (reprinted from Zhejiang Daily). http://www.zj.xinhuanet.com/20250718/144633e1c1434db58bff08fa071ed653/c.html (accessed July 2026).

[8] Liu, X. (2026-01-19). From Virtual Intelligence to Physical Execution: Physical AI — Another Highlight of Artificial Intelligence Development. Science and Technology Daily, Page 4. https://www.stdaily.com/web/gdxw/2026-01/16/content_462757.html (accessed July 2026).

[9] New Zhiyuan. (2026-05-19). Figure Robot Live Streams for 100 Hours, Sorting 130,000 Packages Non-Stop at 3 Seconds Each! Reprinted by Beijing Academy of Artificial Intelligence (BAAI). https://hub.baai.ac.cn/view/54787 (accessed July 2026).

[10] Adcock, B. [@adcock_brett]. (2026, June 30). Figure 03 humanoid robot has started performing a logistics workflow at BMW Group Plant Spartanburg [Post on X]. X. https://x.com/adcock_brett/status/2071973594268327993.

[11] Sun, X. C. (2026-06-24). Zhiyuan Robotics' Peng Zhihui: After Reaching 10,000 Units in Mass Production, a New Digital Milestone Will Be Revealed Soon. Shanghai Securities News · China Securities Network. https://www.cnstock.com/commonDetail/733753 (accessed July 2026).

[12] NVIDIA. (2026-06-01). Develop Physical AI Reasoning, World, and Action Models with NVIDIA Cosmos 3 [NVIDIA Technical Blog]. https://developer.nvidia.com/blog/develop-physical-ai-reasoning-world-and-action-models-with-nvidia-cosmos-3 (accessed July 2026).

[13] NVIDIA. (2026-01-07). Building Generalist Humanoid Capabilities with NVIDIA Isaac GR00T N1.6 Using a Sim-to-Real Workflow [NVIDIA Technical Blog].

https://developer.nvidia.com/blog/building-generalist-humanoid-capabilities-with-nvidia-isaac-gr00t-n1-6-using-a-sim-to-real-workflow/ (accessed July 2026).

[14] Gemini Robotics Team. (2025). Gemini Robotics 1.5: Pushing the Frontier of Generalist Robots with Advanced Embodied Reasoning, Thinking, and Motion Transfer. arXiv, arXiv:2510.03342. https://arxiv.org/abs/2510.03342 (accessed July 2026).

[15] Li, J. H., Zhao, X. J., & Fu, Z. (2016). Research on Computationalism and Its Theoretical Difficulties. Beijing: China Social Sciences Press.

[16] Huo, S. Q. (2019). The Significance and Challenge of Research on the Symbol Grounding Problem in Artificial Intelligence. Journal of Shanghai Normal University (Philosophy & Social Sciences Edition), (3), 98-107. DOI: 10.13852/J.CNKI.JSHNU.2019.03.010.

[17] Zhang, Q. Q. (2018). The Ontology of Meaning: The Origin and Essentials of Philosophical Hermeneutics. Beijing: The Commercial Press.