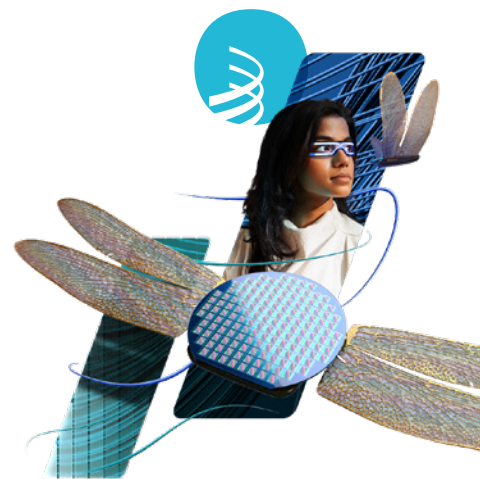


Spirit-AI: science-driven entrepreneurship in embodied AI

Fengtao Han, Founder and CEO, Spirit-AI
Junliang Guo, Head of Foundation Models, Spirit-AI



This paper examines the evolving landscape of robotics, characterized by a transition from deterministic programmed automation toward learning-based embodied artificial intelligence (AI). Our analysis suggests that the primary constraint in contemporary robotics is not mechanical precision, but the development of cross-domain cognitive capabilities, a framework that Spirit-AI refers to as a “universal brain.” This study details the transition of such a system from a laboratory hypothesis to an industrial application.

The findings presented here are based on a “dirty data” strategy, which prioritizes real-world environmental complexity over curated data sets to enhance model resilience. We provide empirical validation for an embodied scaling law, defined by the power-law relationship $F(D) = C \times D^{(-\alpha)}$, which indicates a predictable increase in task efficiency as pre-training data volume scales.

Industrial validation of this framework was conducted through the deployment of the Moz1 platform at CATL. The results, involving the successful processing of over 1,000 battery units in an unstructured production environment, demonstrate the effectiveness of foundation models in handling high-entropy industrial tasks. Finally, this paper discusses the “10-10 mission,” the objective of achieving 10 percent global robot ownership within a decade, as a benchmark for assessing the long-term scalability and economic viability of general-purpose embodied AI.

1. Introduction: the convergence of industrial experience and AI research

1.1 The myth of the university incubator

In the conventional narrative of high-tech innovation, breakthrough robotics ventures are often depicted as direct spin-offs from academic laboratories, as linear translations of PhD theses into commercial prototypes. However, the genesis of Spirit-AI followed a distinct, more strategic trajectory rooted in science-based entrepreneurship.

Rather than emerging from a single university incubator, Spirit-AI was born at the intersection of two critical industry requirements: the robotics sector’s need for an adaptable “brain” and the AI research community’s need for “physical grounding.” Our organizational origins are a synthesis of over a decade of prior industrial experience in robotics, commercialization and corporate research and development (R&D) in end-to-end foundation models. This combination enabled us to design a commercialization pathway that explicitly moves from scientific proof-of-concept (PoC) to industrial pilot testing, and finally to large-scale deployment. By integrating industrial pragmatism with frontier AI research, we sought to reduce traditional “lab-to-market” friction, treating the development of embodied AI as a continuous engineering and scientific loop.

1.2 Resonance: the meeting of two domains

Our inception in 2023 was the result of an alignment between two analytical perspectives on the future of physical intelligence. One perspective was shaped by a comprehensive search for a solution to the industry's "generalization bottleneck," involving deep-dive analysis and technical interviews with over 100 potential partners across the robotics ecosystem. Simultaneously, we were investigating the theoretical limits of embodied AI through the lens of large language models (LLMs), seeking a high-stakes, real-world industrial platform to anchor our hypotheses.

The convergence of these paths led to a fundamental analytical insight: the historical constraints or "ceiling" of the robotics industry were primarily defined by the limitations of rule-based control systems. The rapid evolution of reasoning capabilities in foundation models suggested that this ceiling could be overcome if such models were anchored in physical reality through a universal brain. This thesis suggests that the transformative potential of LLMs is not fully realized in digital spaces, but in their ability to govern complex interactions in the physical world.

1.3 Redefining the problem: from precision to generalization

For the past half-century, the robotics industry has prioritized hardware precision, speed and repeatability. This paradigm succeeded in structured environments, such as automotive assembly lines, but remains insufficient for the unstructured complexity of the broader world.

Our analysis shifts the focus from mechanical precision to cognitive generalization. This shift is supported by evidence from our deployments, which indicates that the missing link in robotics is a system capable of perception, reasoning and action within a unified framework. By treating robotics as a data-driven learning problem rather than a deterministic engineering task, we transition embodied AI from laboratory demonstrations to a potentially scalable infrastructure for general-purpose physical intelligence.

2. Strategic pivots: data velocity over hardware perfection

2.1 The "cognitive friction" in resource allocation

In the early stages of Spirit-AI, we encountered a defining moment of "cognitive friction." This represented a fundamental tension between two paradigms: traditional robotics logic and the AI foundation-model paradigm (Generalist, 2025; Sunday Robotics, 2025). The hardware team, rooted in decades of industrial experience, initially advocated for a sequential development path: they argued for delaying large-scale data harvesting until our proprietary, high-precision and low-cost actuators were finalized. Their concern was "data purity": the fear that training a model on inconsistent, third-party hardware would introduce unmanageable "noise" and hinder the model's ability to achieve high-precision control.

2.2 The pivot: data velocity as the decisive variable

Inspired by Hu's insights into the convergence of vision-language-action (VLA) models (Hu et al., 2024), we made a strategic pivot. We recognized that in the era of foundation models, data velocity, the speed at which a model can encounter and learn from diverse physical scenarios, is a decisive variable. While motors and actuators can be iterated and replaced, the "lost time" in data accumulation is an irrecoverable cost.

By pivoting to a data-first strategy, we utilized existing third-party platforms to begin high-frequency harvesting immediately. This approach was validated through the development of our fourth-generation wearable data collection system. In internal benchmarks compared against traditional teleoperation setups (the baseline), this system reduced data acquisition costs by 90 percent by eliminating the need for specialized, fixed-site infrastructure and reducing work-hours per successful demonstration. (Spirit-AI internal validation data, 2025–2026)

Through internal validation across a wide range of diverse manipulation trajectories, we demonstrated that through algorithm compensation, a sufficiently large model can "absorb"

hardware noise. This method reportedly achieved a 95 percent data usability rate – meaning that data collected from uncalibrated third-party hardware could be successfully integrated into the pre-training set without degrading downstream task performance. This finding, based on Spirit-AI internal validation against a baseline of traditional teleoperation setups, suggests that the model's scale effectively compensates for mechanical variances. This decision shifted our hardware from a bottleneck into a flexible vessel for an ever-evolving universal brain. (Spirit-AI internal validation data, 2025–2026.)

2.3 Algorithm compensation and the “dirty data” strategy

To resolve the noise issue, we implemented algorithm compensation mechanisms. We realized that “clean,” hyper-curated laboratory data are the enemy of robustness. If a model only learns in a perfect hardware environment, it will fail in the “unstructured reality” of a live factory.

By training on what we call “dirty data” defined as data sets containing inherent environmental variance, irregular sensor drift and mechanical tolerances, the model learns the underlying physics of the world rather than specific hardware quirks. Our findings suggest that exposure to these stochastic variables during pre-training is essential for developing adaptive behaviors. This strategy allowed us to scale high-frequency data harvesting across multiple third-party robot platforms, ensuring that the universal brain is grounded in the functional entropy of the real world.

3. Science-driven organization: integration and strategic trade-offs

3.1 The three-pillar co-evolution model

In Spirit-AI, we recognized early on that the complexity of embodied AI demands an organizational structure that mirrors our scientific philosophy. We moved away from the traditional siloing of hardware and software, establishing a three-pillar structure designed for rapid co-evolution:

- **Hardware team:** rather than pursuing mechanical complexity for its own sake, our hardware team focuses on developing the Moz1 platform with a specific emphasis on “learnability.” The hardware is designed hand-in-hand with our model’s requirements to ensure clean data collection and stable locomotion.
- **AI/VLA research team:** this team advances our frontier VLA models. By training exclusively on real-world data, we ensure our models possess the long-horizon planning and high-precision physical skills required for diverse tasks.
- **Application and deployment team:** this team serves as our strategic bridge to physical reality, responsible for fine-tuning behavioral parameters and rigorously validating the generalization capabilities of our models. By ensuring high-success execution across diverse environments ranging from domestic settings to industrial floors, they effectively close the R&D loop.

3.2 Why integration matters

This integrated structure enables rapid iteration and co-evolution. When our application team discovers a limitation in manipulation precision, the hardware team can adjust end-effector design, while the AI team simultaneously works on improving grasp prediction. This tight feedback loop has proven essential for progress.

3.3 What worked: three critical success factors

3.3.1 Investing in premium, learnable hardware

Early in our journey, we made the strategic decision to invest heavily in premium robotics hardware. This was not a luxury; it was a scientific necessity. Robots must move with stability, consistency and repeatability, or the resulting data set becomes noisy and less actionable.

We learned that hardware designed for “learnability” differs fundamentally from hardware optimized for teleoperation or scripted tasks. Every action must be smooth and deterministic, suitable for model training. This upfront investment produced data sets with a superior signal-to-noise ratio, resulting in significantly better generalization and task success.

3.3.2 Collecting diverse, high-quality real-world data

Our most important finding concerns data diversity. Following principles highlighted in Hu's 2024 paper (Hu et al., 2024), we implemented a paradigm called "open-ended, goal-driven data collection." Instead of scripting demonstrations, we instruct data collectors with a simple principle: Do something useful." Our experiments validated this approach through a comparative study between models pre-trained on "goal-driven" data versus those trained on "scripted" task-specific data. In a controlled test involving 15 novel manipulation tasks ranging from "clearing a dining table" to "organizing a toolkit," the diverse-data model reached equivalent success rate performance with 40 percent fewer fine-tuning iterations. This efficiency gain was observed across a sample size of 2,000 fine-tuning trajectories, proving that broad pre-training significantly lowers the barrier for downstream task adaptation. (Spirit-AI internal validation data, 2025–2026.)

3.3.3 Building efficient teleoperation and wearable systems

To enable scalable data collection, we developed efficient teleoperation and wearable systems enabling dexterous demonstrations and natural human-like behaviors. Our data collection methodology includes:

- high-accuracy virtual reality teleoperation for precision manipulation;
- wearable tools for efficient dexterous table-top demonstrations;
- mobile wearable devices enabling scalable collection of natural daily tasks.

These systems, combined with our Spirit Data Hub (Quanta) platform for automated ingestion, quality control, labeling and model training, have enabled us to collect thousands of hours of high-quality real-world data.

3.4 What did not work: the failure of simulation and clean data

3.4.1 Simulated data did not transfer effectively

Despite widespread interest in large-scale simulation, we consistently observed that simulated behaviors lacked the fidelity, noise profile and affordance realism required for high-precision physical execution. The "sim-to-real gap" is not merely a technical challenge to be solved—it reflects a fundamental mismatch between simulation and physical reality. This reinforced our core thesis: real-world embodiment is irreplaceable.

3.4.2 Clean, curated data limits generalization

The robotics field has historically relied on "clean" data sets such as Open X-Embodiment (OXE) (Open X-Embodiment Collaboration, 2024), AgiBot and RoboCOIN (Bu et al., 2025; Wu et al., 2025). These data sets feature carefully designed tasks, controlled object placement and simplified movements to ensure high success rates.

However, we discovered this approach has critical limitations:

- **Low diversity:** by forcing collectors to follow curated scripts, data inherently lack diversity. A "wipe table" data set always involves a table and wiping motion, disconnected from other tasks. The model never learns to recover from slips, handle cluttered surfaces, or adapt to lighting changes.
- **Limited scalability:** each demo-driven data set requires careful task design and quality control. When research effort is a fixed resource, allocating significant portions to designing and controlling tasks inevitably reduces data quantity.

Our experiments suggested that models trained on diverse, "messy" data significantly outperformed those trained on clean, curated data sets when evaluated on novel tasks.

4. Scaling laws: the path to more generalizable robots

4.1 Discovery of the scaling law

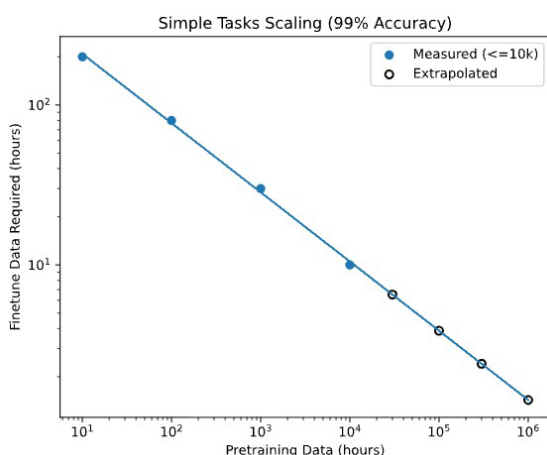
Through systematic experimentation based on Spirit-AI internal validation data, we observed a scaling relationship in embodied AI: as pre-training data volume increases, the efficiency of downstream task adaptation appears to improve predictably. This allows us to model performance as a function of data investment, subject to further external validation.

We fitted this relationship using the following power-law formula:

$$F(D) = C \times D^{-\alpha}$$

where D represents pre-training data size, $F(D)$ represents downstream task accuracy, C is a constant and α is the fitted scaling coefficient.

Figure 1 Simple tasks scaling



Source: Author's own compilation.

4.2 Simple tasks: 30-task validation

On 30 simple tasks (factory battery insertion, pick-and-place, etc.) with a target accuracy of 99 percent (Figure 1), we observed:

- From 10 hours to 10,000 hours of pre-training data, fine-tuning requirements dropped from ~100 hours to ~10 hours per task (89 percent reduction).
- Extrapolating to 1 million hours of pre-training, simple tasks require only minutes of fine-tuning data.

(Spirit-AI internal validation data, 2025–2026.)

The fitted scaling law: $\alpha \approx 0.43$, meaning every 10× increase in pre-training data reduces fine-tuning requirements by ~2.7×.

4.3 Complex tasks: 10 high-difficulty scenarios

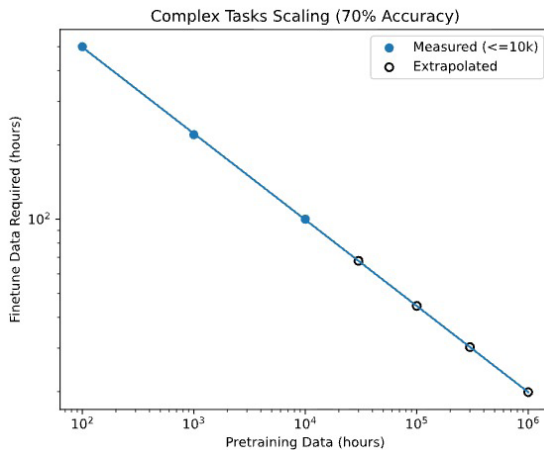
On 10 complex tasks (folding clothes, toy organization, etc.) with 70 percent accuracy targets (Figure 2), we found:

- Complex tasks exhibit similar scaling behavior, albeit with higher absolute data requirements.
- At 1 million hours pre-training, complex tasks require ~29 hours of fine-tuning data.

(Spirit-AI internal validation data, 2025–2026.)

The fitted scaling law: $\alpha \approx 0.35$, meaning every 10× increase reduces fine-tuning by ~2.24×.

Figure 2 Complex tasks scaling



Source: Author's own compilation.

4.4 Implications for the industry

These scaling laws suggest a clear path to more generalizable robots: massive pre-training on diverse real-world data, followed by minimal task-specific fine-tuning.

If pre-training reaches million-hour scales, even complex household tasks may become “few-shot” learning problems.

5. Validation: benchmark results

5.1 RoboChallenge Table30 performance

Our newest model, Spirit-v1.5 (Spirit-AI Robotics, 2026), was evaluated on RoboChallenge Table30 (RoboChallenge Team, 2025), a recently established real-world robot benchmark consisting of 30 diverse manipulation tasks including “arrange_flowers,” “make_vegetarian_sandwich” and “plug_in_network_cable.”

As of January 11, 2026, Spirit-v1.5 achieved leading performance on RoboChallenge Table30 according to the benchmark information cited by the authors. Where the ranking or statistical significance is used as evidence, the authors should ensure that the underlying benchmark record and comparison method are fully traceable (RoboChallenge Team, 2025; Spirit-AI Team, 2026).

5.2 Open-source commitment

To support the research community and foster further innovation in physical intelligence, we have open-sourced Spirit-v1.5:

- Code: <https://github.com/Spirit-AI-Team/spirit-v1.5>
- Model: <https://huggingface.co/Spirit-AI-robotics/Spirit-v1.5>

6. Industrial validation at CATL

6.1 Escaping the “demo trap”

Scientific entrepreneurship requires moving beyond the “demo trap”: the industry-wide tendency to showcase highly scripted, impressive robotic movements in controlled laboratory settings that fail in the messy reality of a factory. At CATL, Spirit-AI’s strategy was defined by “scientific filtration.” We did not seek the most “visually spectacular” task, but the one that would provide the most rigorous validation of our model’s resilience.

6.2 The 1-in-200 process selection

We analyzed over 200 distinct industrial processes, filtering them through a three-dimensional scientific framework:

- **Takt time and inference latency:** we evaluated whether the industrial cycle time allowed for the high-frequency inference required by end-to-end VLA models.
- **Error tolerance and self-correction:** we sought tasks where a “universal brain” could demonstrate its closed-loop control by correcting minor deviations in real time without halting the assembly line.
- **Environmental entropy:** we looked for tasks with high visual and tactile diversity scenarios where traditional, rule-based robots fail due to subtle variations in lighting, material texture, or object positioning.

6.3 Case study: industrial validation in battery manufacturing

While laboratory benchmarks are essential, the true test of an embodied AI system lies in its ability to handle “unstructured reality” on a live production line. We conducted an extensive industrial pilot with a global leader in lithium-ion battery manufacturing, deploying the Moz1 platform for high-precision battery cell inspection and insertion tasks.

In this deployment, we moved beyond “scripted” automation to evaluate the model’s self-correction capabilities under variable industrial conditions. Over a period of several months, the system completed over 1,000 cycles without manual teleoperative intervention. We define “zero intervention” as the model’s ability to autonomously adjust its trajectory in response to mechanical tolerances and sensor drift through visual feedback. This validation suggests that a VLA model trained on diverse “dirty data” can achieve industrial-grade reliability in high-entropy environments where traditional rule-based robotics typically fail.

7. Policy relevance: infrastructure for generalization

To bridge the “generalization bottleneck” at a systemic level, the Spirit-AI case suggests several possible ecosystem lessons:

- **Establishing “embodied AI data gyms”:** these could function as shared testbeds and open foundries for physical data. By providing standardized environments for diverse data harvesting, ecosystem actors could help small-to-medium enterprises reduce data acquisition costs, a significant barrier which Spirit-AI has cut by 90% thanks to its own infrastructure
- **Incentivizing cross-industry data standards:** ecosystems could encourage the creation of “open-ended goal-driven” data sets. This could help move the industry away from siloed, task-specific data toward a shared pool of “physical common sense” that benefits the entire ecosystem.

8. Conclusion: from laboratory hypotheses to social reality

The evolution of Spirit-AI suggests that the primary constraint in robotics today is not hardware, but the cognitive ability to generalize. Our internal scaling-law analysis suggests a possible roadmap: as pre-training data scales, the marginal cost of adapting to new tasks could fall substantially.

We believe embodied AI is in the early stages of a potential transformative breakthrough. Our “10-10 mission” to enable 10 percent of the world to own robots within a decade is a company vision guided by the belief that intelligent machines should be accessible and useful in everyday life. By anchoring AI in the complexity of the physical world, we transition from laboratory demonstrations to a future of functional, scalable and general-purpose embodied AI.

Acknowledgments

The evolution of Spirit-AI is a collective achievement of industrial vision and scientific persistence. We extend our sincere gratitude to Shunwei Capital for their unwavering support of long-term scientific exploration over short-term commercial shortcuts. Their partnership has been instrumental in transitioning our universal brain from a laboratory hypothesis to a scalable industrial reality.

We are also indebted to our frontline engineering teams, whose dedication to harvesting “dirty data” across thousands of steps on factory floors provided the empirical foundation for our scaling laws. Finally, we thank CATL and our strategic partners for their openness to innovation, allowing us to bridge the gap between “demo” and “deployment” on live production lines.

References

Bu, Q., *et al.* (2025). AgiBot World Colosseo: A large-scale manipulation platform for scalable and intelligent embodied systems. arXiv:2503.06669.

Generalist. (2025). GEN-0: Embodied foundation models that scale with physical interaction. Generalist AI Blog, November 4. Available at: <https://generalistai.com/blog/nov-04-2025-GEN-0>.

Hu, Y., *et al.* (2024). Data scaling laws in imitation learning for robotic manipulation. *arXiv preprint*, arXiv:2410.18647.

Open X-Embodiment Collaboration. (2024). Open X-Embodiment: Robotic learning datasets and RT-X models. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.

RoboChallenge Team. (2025). RoboChallenge: Large-scale real-robot evaluation of embodied policies. Available at: <https://robochallenge.ai>.

Spirit-AI. (2025–2026). Internal validation data on data acquisition, data usability, fine-tuning efficiency, scaling-law analysis, and CATL deployment. Unpublished internal data;

Spirit-AI Robotics. (2026). Spirit-v1.5 model card. Hugging Face. Available at: <https://huggingface.co/Spirit-AI-robotics/Spirit-v1.5>.

Spirit-AI Team. (2026). Spirit-v1.5 code repository. GitHub. Available at: <https://github.com/Spirit-AI-Team/spirit-v1.5>.

Sunday Robotics. (2025). ACT-1: A robot foundation model trained on zero robot data. *Sunday AI Journal*. Available at: <https://www.sunday.ai/journal/no-robot-data>.

Wu, S., *et al.* (2025). RoboCOIN: An open-sourced bimanual robotic data collection for integrated manipulation. arXiv:2511.17441.

Copyright

© WIPO, 2026

Attribution 4.0 International (CC BY 4.0)

The CC license does not apply to non-WIPO
content in this publication.

WIPO Reference No. GII2026/EN/EXTCONT

DOI: 10.34667/tind.59195

Front image: Getty Images, Unsplash.