The Collapse of Scaling Laws and the Quest for Intelligence's "Root"
——From the Misallocation of Trillions in Compute to the Logographic AI Paradigm Response

In July 2026, DeepMind researcher Diogo Almeida dropped a bombshell on the AI community with his blog post "Scaling Laws, Honestly"[6]. A former insider at OpenAI who had worked on large model optimization, he pointed directly to a fatal bug in the 2020 Scaling Laws paper that had defined the entire trajectory of AI development[2].

"The original scaling laws were wrong due to a bug."[6]

This was no hindsight. This was someone who had been there, tearing open the wound of that curve.

I. One Bug, Trillions in Compute Down the Drain

In 2020, OpenAI published the landmark paper "Scaling Laws for Neural Language Models"[2]. The conclusion was simple and seductive: under a fixed compute budget, you should prioritize making the model bigger. The formula stated that the optimal parameter count scales as compute to the 0.73 power.

This statement directly defined GPT-3—175 billion parameters. It told the world: don't ask, just scale parameters; as long as you make the model large enough, miracles will happen through sheer force.

Two years later, DeepMind's Chinchilla paper turned this conclusion upside down[3]. A 70-billion-parameter model, trained on 1.4 trillion tokens—less than half the size of GPT-3, but with over four times the data—comprehensively outperformed the 280-billion-parameter Gopher.

With the same compute budget, one became a "puffy" brute, the other a lean, mean fighter.

But the problem went much deeper. Almeida revealed a more fatal detail: the bug in the Kaplan paper was not a calculation error, but an experimental design flaw[6]. OpenAI's original experiment fed all models—regardless of size—the exact same amount of data (about 130B tokens). Small models were overfed, while large models were severely undernourished. They used a cosine learning rate decay, gradually pushing the learning rate to zero as training approached its end, flattening the curve and making it appear as if "the model had learned all it could"[6].

This was not the model's limit. This was the learning rate artificially cutting off the model's growth.

Even more ironically, the Chinchilla paper that "corrected" OpenAI wasn't clean either. In 2024, Besiroglu et al. discovered another bug in Chinchilla's fitting—the loss scale in the optimizer was set too high, causing premature termination of the fitting process[4]. Pearce and Song further pointed out that Kaplan's parameter counting method (excluding embedding layers) and the small scale of their analysis also led to an overestimation of the coefficients[4][9].

The so-called Scaling Laws were never a ironclad physical law like Newton's three laws. They were merely an empirically fitted curve.

II. From "English Scaling Laws" to "Rootless Intelligence"

Deeper questions inevitably followed.

Before diving into the details, Figure 1 visually presents the two layers of limitations in Scaling Laws—from experimental design errors to a paradigm-level language blind spot.

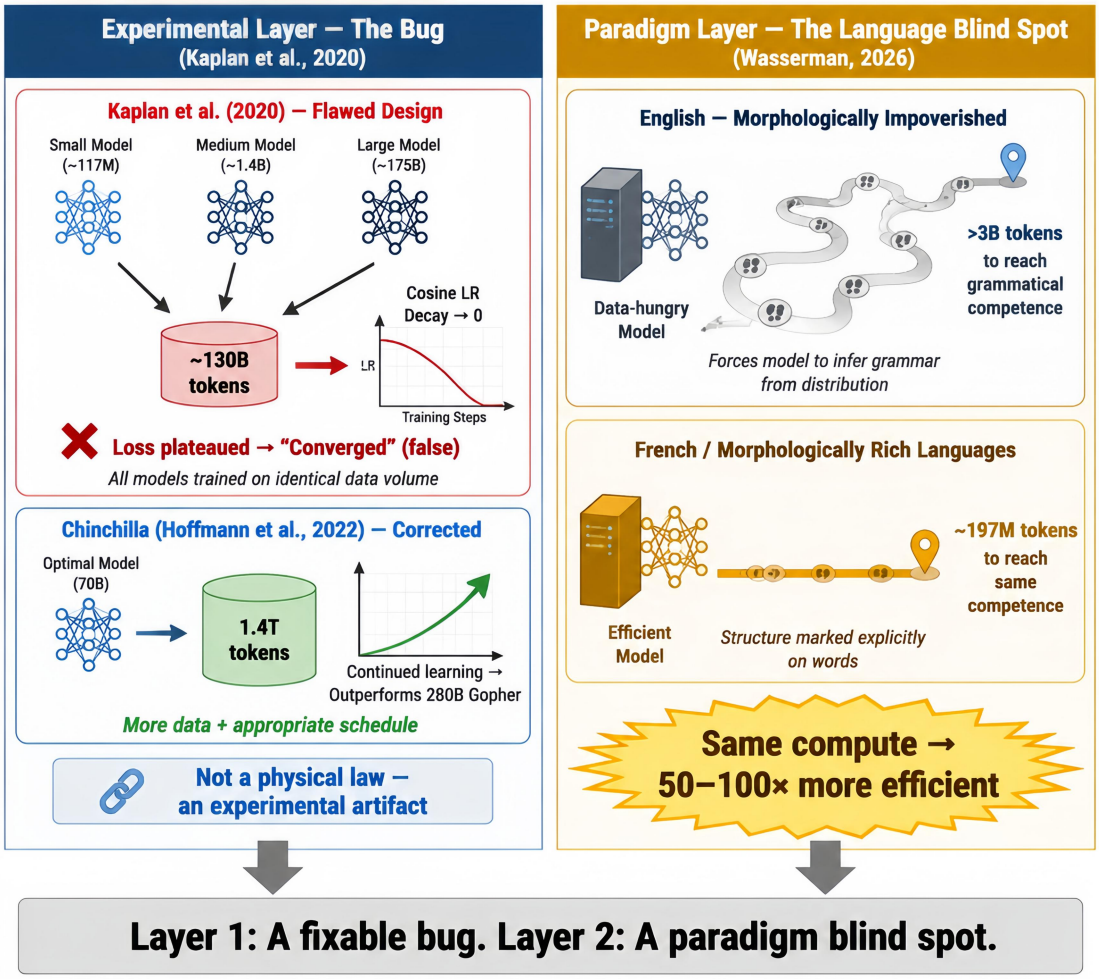

Figure 1. Two Layers of Limitations in Scaling Laws: From Experimental Design Error to Language-Dependent Blind Spot

The left panel shows the experimental design flaw in OpenAI's original Scaling Laws [2]: all models, regardless of size, were trained on the same amount of data (~130B tokens) with a cosine decay learning rate, artificially creating the appearance of convergence. Chinchilla [3] demonstrated that a smaller model (70B) trained on substantially more data (1.4T tokens) achieves superior performance—revealing that the original "convergence" was an artifact of training methodology.

The right panel reveals a deeper, paradigm-level blind spot identified by Wasserman [6]: existing Scaling Laws are, in effect, "English Scaling Laws." English, as a morphologically impoverished language, forces models to infer grammatical patterns from vast amounts of distributional data. Morphologically richer languages like French encode more structural information directly in the words themselves, reaching equivalent grammatical competence with 50–100× greater efficiency on the same compute. This suggests that what has been framed as a universal law of intelligence scaling is, at least in part, a measure of English's inefficiency as a training language—It is not a

self-evident physical law, but rather, to a large extent, an efficiency curve for a specific language (English) under specific experimental conditions.

In the comments section of Almeida's blog, researcher Adam Zachary Wasserman pointed out a blind spot that everyone had overlooked: even with the formula corrected, the current Scaling Laws are merely "English Scaling Laws"[6].

Using the same architecture and the same compute, he found that a French model reached grammatical competence 50 to 100 times more efficiently than an English model[6]. Why? Because English is a "morphologically impoverished" language—it relies too heavily on distributional patterns, forcing the model to infer meaning from massive amounts of data. Whereas languages like French and Chinese, which have higher information density at the morphological or syntactic level, carry substantial explicit information within the words themselves.

The research of linguists like Drożdż et al. provides quantitative evidence for this: statistical analysis of English and Polish texts revealed that Polish (morphologically richer) has a significantly lower Zipf distribution exponent than English[7]. This means that, for texts of the same length, morphologically richer languages carry more structural information. Wasserman's experiment is the AI-era counterpart of this discovery—he trained models with the same compute and found French to be 50 to 100 times more efficient at reaching grammatical competence than English[6].

The subversive implication of this discovery is this: if Scaling Laws are "English-only," then all the compute planning, model design, and even AGI timelines of the past five years have been built on a language-specific empirical formula. This is not a "fixable" bug, but a paradigm-level blind spot—it means the entire industry may have been optimizing on the wrong scale for five years.

In other words, the "intelligence curve" we have been fitting with trillions in compute has largely been compensating for the informational deficiencies of English as a "morphologically impoverished" language. If the original Scaling Laws experiments had been based on Finnish or Chinese, the shape of that curve might have been entirely different. The so-called "universal law" of intelligence is, in reality, just "English's efficiency ceiling."

When you thought you were exploring the physical laws of "general intelligence," you were actually just measuring "how much compute English wastes."

This is another expression of the "rootless semioticism" dilemma identified by Logographic AI (LAI) theory. In The Rooted Intelligence: The Paradigm Revolution of Logographic AI, Liu Shen diagnoses the philosophical essence of the current mainstream AI paradigm as "rootless semioticism"—where the meaning of symbols is entirely defined by statistical co-occurrence within the symbolic system itself, failing to reach the "referent" beyond the symbol—namely, the real world, human experience, and civilizational values[1].

Token-based AI is, in essence, "wandering" in statistical space—it knows "what," but not "why."

III. AGI Is Not About Being Bigger, But About Being Smarter

The collapse of Scaling Laws does not signal the end of AI, but an opportunity for paradigm shift.

Chinese Academy of Engineering academician Li Guojie, in his paper "Historic Breakthroughs and Great Challenges in Intelligent Computing Technology," explicitly stated that scaling laws may have hit a ceiling[8]. He cites Richard Sutton's judgment—that achieving AGI through large language models "has no future"—and expresses agreement with LeCun's vision of world models. This aligns with Ilya Sutskever's earlier expressions of concern—he noted that "the usable new data in this industry is approaching depletion." In November 2025, he further declared in an interview: "We are moving from the scaling era to the research era."

Peking University researchers Deng Xiaotie and Li Hanyu, in their paper "How Large Language Models Need Symbolism" published in the National Science Review, systematically argue for this shift[5]. They point out that while Scaling Laws are effective in data-rich domains, in the areas of logical reasoning and abstraction that human reasoning depends on—where available data is inherently scarce—AI must return to the power of symbols. Leibniz's calculus notation (dy/dx), they note, surpassed Newton's dot notation precisely because symbols themselves are a form of "cognitive technology"—they allow one to "guess" rules rather than merely memorize them[5].

When "bigger" is no longer the answer, "smarter" becomes the only direction.

IV. Logographic AI's Answer: Let Cognitive Primitives Carry Meaning Intrinsically

From the perspective of Logographic AI, the collapse of Scaling Laws and the revelation of "English Scaling Laws" precisely confirm the structural limitations of the Tokenist paradigm.

Logographic AI replaces the Token with the "Morpho-Root" as the cognitive primitive. Each Morpho-Root is a structured triple[1]:

r = ⟨ S, A, R⟩

- **S (Symbol)** : The symbolic identifier
- **A (Attributes)** : The attribute set, embedding the Morpho-Root's intrinsic semantic features and value constraints
- **R (Relation Functions)** : The relation function set, defining the presupposed logical connections between this Morpho-Root and others

The key design of the Morpho-Root is that meaning is not "emerged" from statistics, but is pre-configured as an intrinsic attribute of the cognitive primitive[1].

When researchers discover that French models achieve grammatical competence 50 to 100 times more efficiently than their English counterparts, what they are truly revealing is this: **the structural information inherent in language itself can replace massive amounts of statistical fitting**. This is the starting point of Morpho-Root theory—the Morpho-Root carries the "innate information" of linguistic structure, rather than being a statistical pattern that must be "guessed" from massive amounts of Tokens.

Wasserman's experiment, at the operational level, validates the core hypothesis of Natural Language Ontology (NLO): cognitive efficiency depends not on the scale of data fitting, but on the information density of primitives (morphemes/characters). While Tokenism must rely on hundreds

of billions of parameters to compensate for English's informational deficiencies, Logographic AI's "Morpho-Root" reduces this need for "statistical compensation" at the architectural level by pre-encoding semantic attribute sets—this explains why structurally rich languages can naturally achieve efficiency gains without additional compute. In other words, Logographic AI does not "correct the slope" on the English curve; rather, it replaces the horizontal axis of the entire coordinate system from "parameter scale" to "structural depth."

The philosophical foundation of Logographic AI is **Natural Language Ontology (NLO)** : language structure itself is a form of "ontology" of intelligence—a primordial way of cognizing the world. Form is meaning—in logographic writing systems, the formal structure of symbols is itself the source of meaning.

V. Conclusion: From Statistical Fitting to Structural Understanding

What Diogo Almeida tore open was not just a bug in a paper, but the illusion of an era[6].

Of course, not everyone agrees that Scaling Laws have completely failed. After declaring "the scaling era is over," Ilya Sutskever also emphasized that scaling itself has not completely stalled—what is truly missing are structural cognitive mechanisms like "value functions." The Deng Xiaotie team points out that while Scaling Laws are still effective in data-rich domains, in abstract domains requiring logical reasoning, AI must return to the power of symbols[5].

These different directions of exploration—Ilya's "value functions," the Deng team's "symbolization," Wasserman's revelation of "language efficiency"—all point to the same direction: **from "scale" to "structure."** This is entirely consistent with Logographic AI's Morpho-Root theory.

"Scale" is not a universal formula. "Bigger" is not an escalator to AGI. The curve that was once treated as gospel is essentially an empirically fitted formula, based on English as the "most compute-wasting" language, under limited experimental conditions.

The true breakthrough in intelligence points to a different possibility: let cognitive primitives carry meaning intrinsically, moving reasoning from "statistical guessing" to "structural traversal."

When researchers discover that French models are 50 to 100 times more efficient with the same compute[6], when Academician Li Guojie calls for "more effective, more reliable, and more energy-efficient new approaches"[8], when the Deng Xiaotie team argues for the necessity of "symbols as a cognitive technology"[5]—they all point toward the direction that Logographic AI is pioneering: from rootless statistical species to rooted meaning species.

The future of intelligence lies not in being bigger, but in being smarter. **Intelligence must have roots; civilization must have a soul.**

References

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

[2] Kaplan, J., McCandlish, S., Henighan, T., et al. (2020). Scaling Laws for Neural Language Models. arXiv, arXiv:2001.08361.

[3] Hoffmann, J., Borgeaud, S., Mensch, A., et al. (2022). Training Compute-Optimal Large Language Models. arXiv, arXiv:2203.15556.

[4] Weng, L. (2026, June 24). Scaling Laws, Carefully. Lil'Log. https://lilianweng.github.io/posts/2026-06-24-scaling-laws/

[5] Deng, X., & Li, H. (2025). How Large Language Models Need Symbolism. National Science Review, 12(10), nwaf339. https://doi.org/10.1093/nsr/nwaf339

[6] Almeida, D. (2026). Scaling Laws, Honestly. Complete Skeptic. https://www.completeskeptic.com/p/scaling-laws-honestly

[7] Drożdż, S., Kwapień, J., & Otrczyk, A. (2009). Approaching the linguistic complexity. In Complex Sciences, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (Vol. 4, pp. 1044-1050). Springer. doi:10.1007/978-3-642-02466-5_104 arXiv:0901.3291

[8] Li, G. (2025). Historic Breakthroughs and Great Challenges in Intelligent Computing Technology. Journal of Integration Technology, 14(1), 1-8. DOI: 10.12146/j.issn.2095-3135.20241215001

[9] Pearce, T., & Song, J. (2024). Reconciling Kaplan and Chinchilla Scaling Laws. TMLR 2024. arXiv:2406.12907.