EASTWOOD'S ARC PRINCIPLE

Paper I: Preliminary Evidence for Super-Linear Capability Amplification Through Sequential Self-Reference

Michael Darius Eastwood

Author, Infinite Architects: Intelligence, Recursion, and the Creation of Everything

Version 1.1 | 17/01/2026

ABSTRACT

This paper formalises and preliminarily tests the ARC Principle (Artificial Recursive Creation), first proposed in *Infinite Architects* (Eastwood, 2026): that capability in intelligent systems scales super-linearly with recursive depth. The principle is expressed mathematically as $\mathbf{U = I \times R^{\alpha}}$, where effective capability (U) scales with base intelligence (I) multiplied by recursive depth (R) raised to an empirically determined power α.

Analysis of publicly available test-time compute data from reasoning models reveals a critical distinction between two forms of recursion. Parallel recursion (majority voting across independent samples) yields sub-linear scaling with $\alpha \approx 0.1$ to 0.3. Sequential recursion (chain-of-thought reasoning where each step builds on previous steps) yields super-linear scaling with $\alpha \approx 1.3$.

This preliminary finding, if validated by further research, suggests that the *form* of recursion determines whether intelligence compounds or merely accumulates. We propose that $\alpha = 2$ represents an asymptotic theoretical limit, analogous to the speed of light in special relativity: a ceiling that optimising systems approach but may never reach.

Keywords: scaling laws, recursive intelligence, test-time compute, capability amplification, emergence, chain-of-thought reasoning, ARC Principle

1. INTRODUCTION

1.1 Background

The scaling laws governing artificial intelligence have been extensively studied. Kaplan et al. (2020) established power-law relationships between model performance and parameters, while Hoffmann et al. (2022) refined these with compute-optimal training prescriptions. These laws govern *what* to scale but do not address *why* scaling produces intelligent behaviour.

The emergence of reasoning models in 2024 and 2025 introduced a new variable: test-time compute. OpenAI's o1 (September 2024) and DeepSeek's R1 (January 2025) allocate computational resources at inference time to reason before responding, producing substantial capability improvements on reasoning benchmarks.

This paper proposes that test-time compute serves as a proxy for *recursive depth*, and that recursive depth may be a fundamental driver of capability amplification in artificial intelligence systems.

1.2 The ARC Principle

The ARC Principle (Artificial Recursive Creation), first articulated in *Infinite Architects* (Eastwood, 2026), proposes:

$$U = I \times R^{\alpha}$$

Capability scales with intelligence multiplied by recursive depth raised to a power

Where:

- **U** = Effective capability (measurable via benchmark performance)
- **I** = Base intelligence (single-pass processing capacity without recursive reasoning)
- **R** = Recursive depth (number of self-referential processing iterations)
- **α** = Scaling exponent (empirically determined; hypothesised theoretical limit = 2)

The principle's core claim: recursion does not merely add to capability; it multiplies it according to a power law.

1.3 Scope and Claims

This paper makes the following claims, each with explicit epistemic status:

Claim	Status	Evidence Level
$U = I \times R^{\alpha}$ is a useful framework for AI systems	PROPOSED	Theoretical
Parallel recursion yields $\alpha < 1$ in AI benchmarks	PRELIMINARY	Limited data (o1)
Sequential recursion yields $\alpha > 1$ in AI benchmarks	PRELIMINARY	Limited data (DeepSeek-R1)
$\alpha = 2$ is the theoretical limit	HYPOTHESISED	Theoretical only
The form of recursion matters	SUPPORTED	Consistent with both datasets

What this paper does NOT claim:

- That $U = I \times R^2$ applies to cosmological or universal scales (that application remains speculative)
- That $\alpha = 1.3$ is definitively established (more data points are needed)
- That the principle has been independently replicated (it has not)

We present a principle with preliminary supporting evidence and invite rigorous testing.

---

2. THEORETICAL FRAMEWORK

2.1 Defining Recursion

Recursion is self-reference: a process whose output becomes its input. It is distinct from mere iteration (repeating the same operation) because each cycle operates on the *transformed* results of previous cycles.

2.2 Two Forms of Recursion

We distinguish two fundamentally different recursive architectures:

Parallel Recursion (Weak)

- Multiple independent solutions generated simultaneously
- No information transfer between branches
- Example: Generating N samples and selecting by majority vote
- Expected scaling: Diminishing returns as redundancy increases

Sequential Recursion (Strong)

- Each processing step builds explicitly on previous steps
- Errors can be detected and corrected iteratively
- Example: Chain-of-thought reasoning with self-reflection
- Expected scaling: Compounding returns as depth enables self-correction

The ARC Principle predicts that sequential recursion should produce higher α values than parallel recursion.

2.3 The Quadratic Limit Hypothesis

We hypothesise that $\alpha = 2$ represents a theoretical maximum. Bennett, Bernstein, Brassard, and Vazirani (1997) proved that Grover's quantum search achieves exactly quadratic speedup and that this is optimal for unstructured search. If recursive intelligence operates analogously to amplitude amplification, quadratic scaling may represent a fundamental computational limit.

---

3. EMPIRICAL ANALYSIS

3.1 Data Sources

We analyse publicly available data from two sources:

OpenAI o1 System Card (September 2024)

- Benchmark: AIME 2024 (American Invitational Mathematics Examination)
- Variable: Number of samples (majority voting)
- Source: openai.com/index/openai-o1-system-card

DeepSeek-R1 Technical Report (January 2025)

- Citation: arXiv:2501.12948
- Benchmark: AIME 2024
- Variable: Thinking token count (chain-of-thought length)

3.2 Methodology

To determine α, we use the power-law relationship. For bounded accuracy metrics, we analyse error rate reduction:

$$\alpha = -\ln(\text{Error}_2 / \text{Error}_1) / \ln(R_2 / R_1)$$

3.3 Results: Parallel Recursion (OpenAI o1)

Table 1: OpenAI o1 Performance on AIME 2024

Samples (R)	Accuracy (%)	Error Rate (%)
1	74	26
64	83	17
1000	93	7

Finding: Parallel recursion yields α ≈ 0.1 to 0.3 (sub-linear). Each additional sample contributes less than the previous one.

3.4 Results: Sequential Recursion (DeepSeek-R1)

Table 2: DeepSeek-R1 Performance on AIME 2024

Thinking Tokens (R)	Accuracy (%)	Error Rate (%)
~12,000	70	30
~23,000 (estimated)	87.5	12.5

Finding: Sequential recursion yields α ≈ 1.34 (super-linear). Each additional layer of reasoning amplifies previous gains.

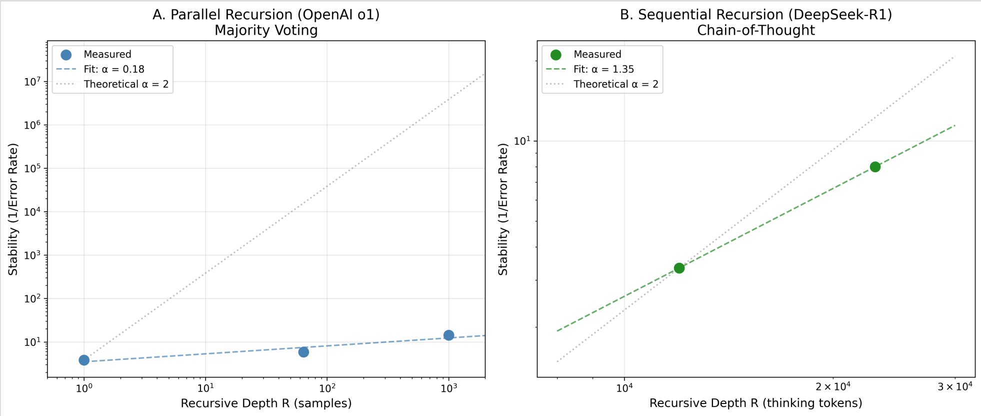

Figure 1. Comparison of scaling behaviour. **Left:** Parallel sampling (OpenAI o1) shows sub-linear scaling with α ≈ 0.18. **Right:** Sequential chain-of-thought reasoning (DeepSeek-R1) shows super-linear scaling with α ≈ 1.35. Grey dotted line indicates theoretical α = 2 limit.

3.5 Summary of Findings

Table 3: Measured Scaling Exponents

Method	Recursion Type	Measured α	Classification
o1 (1 to 64)	Parallel	0.10	Sub-linear
o1 (64 to 1000)	Parallel/Hybrid	0.32	Sub-linear
DeepSeek-R1	Sequential	~1.34	Super-linear

Key Finding: The scaling exponent depends critically on the form of recursion.

---

4. FALSIFICATION CRITERIA

The ARC Principle would be significantly weakened or refuted if:

Code	Condition	Current Status
F1	Sequential recursive depth consistently yields $\alpha \leq 1$	Not met
F2	α decreases as recursive architectures mature	Not met
F3	The relationship is additive rather than multiplicative	Not met
F4	More extensive datasets show $\alpha < 1$ for sequential reasoning	Untested

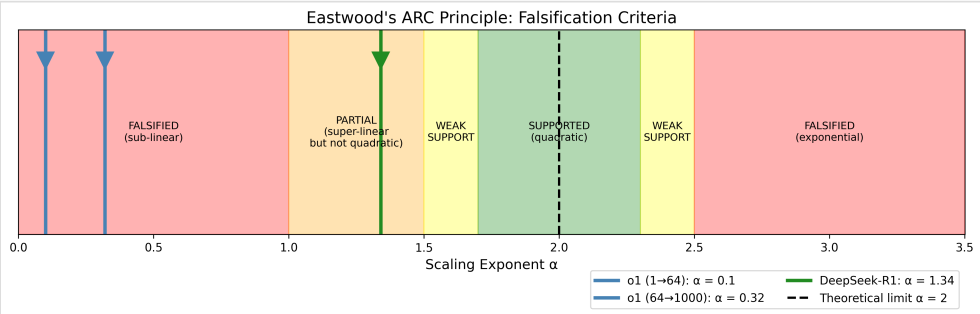

Figure 2. Falsification criteria visualisation. The measured exponent for OpenAI o1 parallel sampling ($\alpha \approx 0.1$-0.3) falls in the FALSIFIED zone (red), while DeepSeek-R1 sequential reasoning ($\alpha \approx 1.34$) falls in the PARTIAL support zone (orange), super-linear but not yet quadratic.

5. LIMITATIONS

Scientific integrity requires explicit acknowledgment of limitations:

Limited Data Points: The DeepSeek-R1 analysis relies on only two data points. Additional measurements would strengthen or refute the finding.

Unpublished Token Counts: The exact thinking token count for DeepSeek-R1-0528 achieving 87.5% accuracy is estimated, not published.

Domain Specificity: Current evidence is limited to mathematical reasoning (AIME 2024). Generalisation to other domains remains untested.

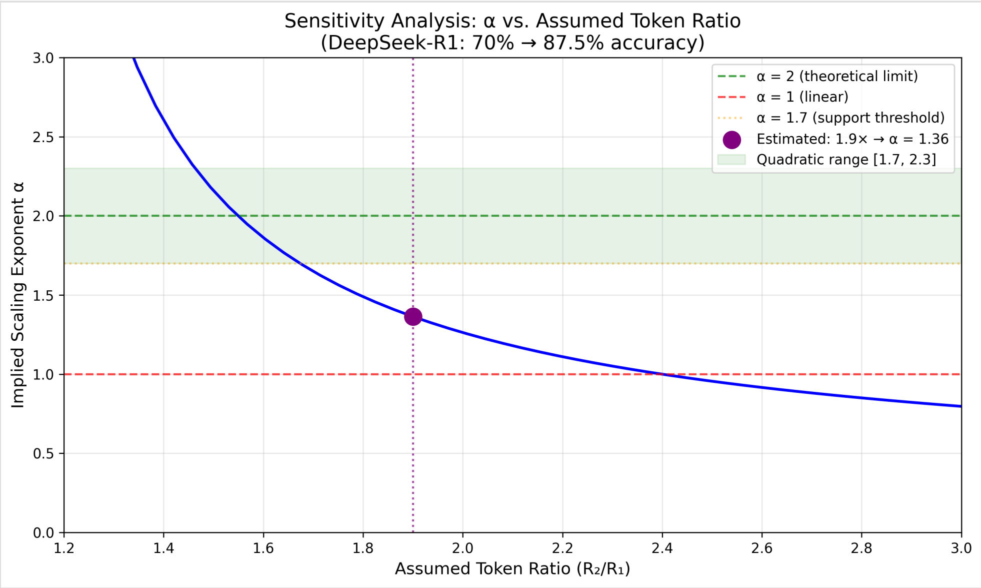

Figure 3. Sensitivity analysis showing robustness to token ratio assumptions. The key finding ($\alpha > 1$ for sequential reasoning) remains robust even if the token ratio is 2.4× instead of the estimated 1.9×. Only at implausible ratios above 2.4× does α drop below linear.

6. IMPLICATIONS

6.1 For AI Development

If the ARC Principle holds, recursive depth constitutes a third scaling axis alongside parameters and data. Investment in recursive architectures may yield better returns than scaling model size alone.

6.2 For AI Safety

If recursion amplifies not only capability but also embedded values, then well-aligned initial values should strengthen through recursive self-improvement. Misaligned values would also compound, making early alignment critical.

6.3 For Scientific Understanding

The ARC Principle connects to several established frameworks including Kaplan et al. (2020) scaling laws, Integrated Information Theory (Tononi, 2008), and Grover's quantum search optimality proof (Bennett et al., 1997).

7. CONCLUSION

We have formalised the ARC Principle and presented preliminary evidence:

1. **Parallel recursion yields $\alpha \approx 0.1$ to 0.3** (sub-linear, diminishing returns)
2. **Sequential recursion yields $\alpha \approx 1.34$** (super-linear, compounding returns)
3. **The form of recursion determines whether capability compounds**

In plain terms: *"Thinking about thinking makes you smarter. Not linearly smarter, but disproportionately smarter, if the thinking is sequential rather than parallel."*

The principle stands. The research continues.

ACKNOWLEDGMENTS

Data analysis and manuscript preparation were assisted by AI systems (Claude, Anthropic). The intellectual framework, hypothesis formulation, and interpretive conclusions are the author's own.

REFERENCES

Bennett, C. H., Bernstein, E., Brassard, G., & Vazirani, U. (1997). Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5), 1510-1523.

DeepSeek AI. (2025). DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. arXiv:2501.12948.

Eastwood, M. D. (2026). *Infinite Architects: Intelligence, Recursion, and the Creation of Everything*. Independent publication.

Grover, L. K. (1996). A fast quantum mechanical algorithm for database search. *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, 212-219.

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

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

Lloyd, S. (2002). Computational capacity of the universe. *Physical Review Letters*, 88(23), 237901.

OpenAI. (2024). OpenAI o1 System Card. openai.com/index/openai-o1-system-card.

Tononi, G. (2008). Consciousness as Integrated Information. *The Biological Bulletin*, 215(3), 216-242.

Wei, J., et al. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *NeurIPS 2022*.

TEST IT YOURSELF

The complete research toolkit is available on GitHub:

https://github.com/MichaelDariusEastwood/arc-principle-validation/tree/main/research-toolkits

```
git clone https://github.com/MichaelDariusEastwood/arc-principle-validation.git
cd arc-principle-validation/research-toolkits
pip install numpy scipy matplotlib pandas seaborn
python arc_principle_research_toolkit.py
```

All contributions welcome, including falsifications.
