EASTWOOD'S ARC PRINCIPLE

Artificial Recursive Creation: Preliminary Evidence for Super-Linear Capability Amplification Through Sequential Self-Reference

Michael Darius Eastwood

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

January 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 **U = I × R^α**, 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 α ≈ 0.1 to 0.3. Sequential recursion (chain-of-thought reasoning where each step builds on previous steps) yields super-linear scaling with α ≈ 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 α = 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. This hypothesis is grounded in Bennett et al.'s (1997) proof that Grover's quantum search algorithm achieves exactly quadratic speedup and that this is optimal for unstructured search.

The ARC Principle, if confirmed, would establish recursive depth as a third fundamental scaling axis for AI development alongside parameters and data.

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 × R^α

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. The exponent α determines whether this multiplication produces sub-linear (α < 1), linear (α = 1), or super-linear (α > 1) returns.

1.3 Scope and Claims

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

Claim	Status	Evidence Level
U = I × R^α is a useful framework for AI systems	PROPOSED	Theoretical
Parallel recursion yields α < 1 in AI benchmarks	PRELIMINARY	Limited data (o1)
Sequential recursion yields α > 1 in AI benchmarks	PRELIMINARY	Limited data (DeepSeek-R1)
α = 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 × R² applies to cosmological or universal scales (that application remains speculative and lacks operationalisation)
- That α = 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.

1.4 Why "Principle" Rather Than "Law"

We designate this a "principle" rather than a "law" for epistemological reasons:

- A **law** (such as Ohm's Law or Hubble's Law) describes an extensively verified empirical regularity
- A **principle** (such as the Heisenberg Uncertainty Principle or Pauli Exclusion Principle) states a foundational constraint from which other results follow

The ARC Principle proposes that recursive self-reference is the mechanism by which capability amplifies, with the squared term representing a theoretical limit. The specific exponent α is a parameter within the principle, not the principle itself. This framing makes the principle robust to refinement as more data becomes available.

Additionally, "ARC Principle" (Artificial Recursive Creation) was articulated in *Infinite Architects* (2026), establishing prior publication.

---

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.

Examples of recursive structures:

- DNA contains instructions for replicating DNA

- Evolution produces organisms that participate in evolution
- Consciousness involves a mind modelling itself
- Chain-of-thought reasoning involves reasoning about reasoning

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. This prediction is testable and, as shown in Section 3, preliminarily supported.

2.3 The Quadratic Limit Hypothesis

We hypothesise that α = 2 represents a theoretical maximum, analogous to the speed of light (c) in special relativity. Several theoretical considerations support this:

Grover's Algorithm (Primary Justification): Quantum search achieves exactly quadratic speedup ($O(\sqrt{N})$ versus $O(N)$) over classical search. Bennett, Bernstein, Brassard, and Vazirani (1997) proved this is optimal for unstructured search. No algorithm can do better. If recursive intelligence operates analogously to amplitude amplification, quadratic scaling may represent a fundamental computational limit.

Network Topology: In a fully connected recursive system of R layers, the number of potential information pathways scales as $R(R-1)/2 \approx R^2$, suggesting quadratic complexity growth.

Geometric Expansion: If each recursive iteration expands the accessible solution space multiplicatively rather than additively, quadratic scaling emerges naturally from the compounding of two-dimensional expansion.

Real systems likely achieve α < 2 due to:

- Noise: Not every processing step achieves genuine self-reference
- Imperfect error correction: Mistakes propagate rather than being eliminated
- Computational overhead: Managing recursive state consumes resources
- Saturation effects: Performance approaches hard limits (e.g., 100% accuracy)

---

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)
- Metric: Accuracy percentage
- 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)
- Metric: Accuracy percentage

Note on DeepSeek data: The DeepSeek-R1-0528 model card documents that accuracy increased from 70% to 87.5% with "enhanced thinking depth" and that the original model used approximately 12,000 tokens on average. The exact token count for the 87.5% result is not published; we estimate approximately 1.9× the original token count based on the documented "enhanced thinking depth." This estimate should be verified against the full technical report when available.

3.2 Methodology

To determine α, we use the power-law relationship:

$U \propto R^\alpha$

For bounded accuracy metrics (maximum 100%), we analyse error rate reduction:

$(Error_2 / Error_1) = (R_2 / R_1)^{(-\alpha)}$

Solving for α:

$\alpha = -\ln(Error_2 / Error_1) / \ln(R_2 / R_1)$

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

Source: OpenAI o1 System Card (2024). Note: The 1000-sample result uses re-ranking with a learned scoring function, not pure majority voting.

Calculation (1 to 64 samples):

```
Error ratio: 17/26 = 0.654
R ratio: 64/1 = 64
α = −ln(0.654) / ln(64) = 0.42 / 4.16 = 0.10
```

Calculation (64 to 1000 samples):

```
Error ratio: 7/17 = 0.412
R ratio: 1000/64 = 15.6
α = −ln(0.412) / ln(15.6) = 0.89 / 2.75 = 0.32
```

Finding: Parallel recursion yields $\alpha \approx 0.1$ to 0.3 (sub-linear). Each additional sample contributes less than the previous one. This is consistent with the theoretical expectation for weak recursion.

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

Source: DeepSeek-R1 Technical Report, arXiv:2501.12948 (2025) and DeepSeek-R1-0528 model card

Calculation:

```
Error ratio: 12.5/30 = 0.417
R ratio: 23,000/12,000 = 1.917
α = −ln(0.417) / ln(1.917) = 0.875 / 0.651 = 1.34
```

Finding: Sequential recursion yields $\alpha \approx 1.34$ (super-linear). Each additional layer of reasoning amplifies previous gains. This is consistent with the theoretical expectation for strong recursion.

Caveat: This calculation relies on an estimated token ratio. The exact value should be verified against published data. However, the qualitative finding ($\alpha > 1$ for sequential reasoning) is robust across reasonable token ratio estimates.

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. Parallel recursion (sampling) produces sub-linear returns ($\alpha < 1$). Sequential recursion (chain-of-thought) produces super-linear returns ($\alpha > 1$).

4. LIMITATIONS

Scientific integrity requires explicit acknowledgment of limitations:

4.1 Limited Data Points

The DeepSeek-R1 analysis relies on only two data points (12,000 and 23,000 tokens). While sufficient to estimate a slope, this is inadequate for confident power-law fitting. Additional data points across a wider range would strengthen or refute the finding.

4.2 Unpublished Token Counts

The exact thinking token count for DeepSeek-R1-0528 achieving 87.5% accuracy is not published in the model card. The $\alpha \approx 1.34$ estimate depends on an assumed token ratio derived from the documented "enhanced thinking depth." This assumption requires verification.

4.3 Proxy Measures

"Thinking tokens" may not perfectly capture recursive depth. Token count is a proxy; true recursion is an architectural property (self-referential processing loops). The relationship between token count and genuine recursive depth remains to be established rigorously.

4.4 Saturation Effects

Accuracy percentages saturate at 100%, potentially distorting power-law fits near upper bounds. The 87.5% accuracy at 23,000 tokens approaches this ceiling.

4.5 Domain Specificity

Current evidence is limited to mathematical reasoning (AIME 2024). Generalisation to other domains (language understanding, visual reasoning, planning) remains untested.

4.6 Alternative Explanations

The observed scaling could reflect:

- Log-linear relationships that approximate power laws over limited ranges
- Task-specific effects unrelated to recursion per se
- Training procedure artefacts rather than fundamental properties

4.7 No Independent Replication

These calculations have not been independently replicated. All researchers are encouraged to verify these findings using the publicly available source data.

---

5. WHAT CAN BE FAIRLY CLAIMED

5.1 Claims with Strong Defensibility

Claim	Defensibility	Justification
The equation form $U = I \times R^\alpha$ is novel	Strong	No prior formulation found in literature
The parallel/sequential distinction with measured exponents is novel	Strong	Original framing and analysis
"Eastwood's ARC Principle" as attribution	Strong	Priority established in *Infinite Architects* (2026)
Grover's bound provides theoretical grounding for $\alpha = 2$	Strong	Bennett et al. (1997) is verified

5.2 Claims with Moderate Defensibility

Claim	Defensibility	Justification
$\alpha \approx 1.34$ for DeepSeek-R1 sequential reasoning	Moderate	Reproducible calculation, but limited data points
Sequential recursion yields $\alpha > 1$	Moderate	Consistent with theory and data, but needs replication
Parallel recursion yields $\alpha < 1$	Moderate	Consistent with theory and data, verified from OpenAI

5.3 Claims with Weak Defensibility

Claim	Defensibility	Justification

α = 2 as theoretical limit	Weak	Hypothesis only, no direct evidence yet
The principle applies beyond AI to biology/cosmology	Weak	Speculative, not operationalised
α will increase as architectures mature	Weak	Prediction, not observation

6. FALSIFICATION CRITERIA

The ARC Principle would be significantly weakened or refuted if:

Code	Condition	Current Status
F1	Sequential recursive depth consistently yields α ≤ 1	Not met
F2	α decreases as recursive architectures mature	Not met
F3	The relationship is additive (I + R) rather than multiplicative (I × R)	Not met
F4	More extensive datasets show α < 1 for sequential reasoning	Untested
F5	Cross-domain measurements show no consistent pattern	Untested

Critical test (F2): If future AI systems with improved recursive architectures show α ≤ 1.3 rather than higher values, the hypothesis that α approaches 2 would be weakened.

7. IMPLICATIONS

7.1 For AI Development

If the ARC Principle holds, recursive depth constitutes a third scaling axis alongside parameters and data. This suggests that investment in recursive architectures (chain-of-thought, self-reflection, iterative refinement) may yield better returns than scaling model size alone, particularly for reasoning-intensive tasks.

The strategic implications are significant: the companies currently developing reasoning models (OpenAI, DeepSeek, Anthropic, Google) may be capturing a qualitatively different form of capability scaling.

7.2 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
- The window for intervention may narrow faster than linear projections suggest

This elevates the importance of alignment research conducted before recursive systems become widespread.

7.3 For Scientific Understanding

The ARC Principle connects to several established frameworks:

- **Kaplan et al. (2020):** Adds recursion axis to parameter/data scaling
- **Integrated Information Theory (Tononi, 2008):** Φ as recursion-dependent measure
- **Lloyd (2002):** Universe as quantum computer; recursive cosmic computation
- **Grover (1996), Bennett et al. (1997):** Quadratic speedup as fundamental computational limit

If recursive amplification is general rather than AI-specific, it may illuminate emergence in biological evolution, neural development, and other complex systems. However, such extensions remain speculative until

operationalised.

---

8. CONCLUSION

8.1 Summary

We have formalised the ARC Principle (Artificial Recursive Creation) and presented preliminary evidence:

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

8.2 The Claim

In plain terms, the ARC Principle proposes:

"Thinking about thinking makes you smarter. Not linearly smarter, but disproportionately smarter, if the thinking is sequential rather than parallel. Each layer of genuine self-reflection amplifies previous gains."

This was intuited philosophically in *Infinite Architects*. It is now preliminarily measured empirically. Rigorous validation requires additional data.

8.3 Future Directions

Validation requires:

- Additional data points across wider ranges of recursive depth
- Controlled experiments varying only recursive architecture
- Cross-domain testing beyond mathematical reasoning
- Independent replication

8.4 Final Word

The ARC Principle proposes that recursive self-reference is a fundamental amplifier of capability, following a measurable power law. The principle may prove incomplete. The exponent may require refinement. But the core insight, that the *form* of recursion determines scaling behaviour, is now supported by preliminary empirical evidence.

The principle stands. The research continues.

ACKNOWLEDGMENTS

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). Learning to Reason with LLMs. openai.com/index/learning-to-reason-with-llms.

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

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

Wei, J., Wang, X., Schuurmans, D., 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/VerdictUK/infinite-architects-website/tree/main/arc-research-toolkit

The toolkit includes:

- Full Python implementation (~800 lines)
- Pre-computed results for verification
- Sensitivity analysis tools
- Publication-quality visualisations
- Functions to add your own data and test new models

```
git clone https://github.com/VerdictUK/infinite-architects-website.git
cd infinite-architects-website/arc-research-toolkit
pip install numpy scipy matplotlib pandas seaborn
python arc_principle_research_toolkit.py
```

All contributions welcome, including falsifications.

APPENDIX A: REPRODUCIBILITY

All calculations can be verified from publicly available sources:

OpenAI o1 Data:

- Source: OpenAI o1 System Card (September 2024)
- URL: openai.com/index/openai-o1-system-card
- Table: Performance on AIME 2024 with varying sample counts

DeepSeek-R1 Data:

- Source: DeepSeek-R1 Technical Report
- Citation: arXiv:2501.12948 (January 2025)
- Additional: DeepSeek-R1-0528 model card on HuggingFace

Calculation Code (Python):

```python
import math

# DeepSeek-R1 calculation
```

```
R1, R2 = 12000, 23000
E1, E2 = 0.30, 0.125

alpha = -math.log(E2/E1) / math.log(R2/R1)
print(f"alpha = {alpha:.2f}")  # Output: alpha = 1.34

# OpenAI o1 calculation (1 to 64 samples)
R1_o1, R2_o1 = 1, 64
E1_o1, E2_o1 = 0.26, 0.17

alpha_o1 = -math.log(E2_o1/E1_o1) / math.log(R2_o1/R1_o1)
print(f"alpha (o1) = {alpha_o1:.2f}")  # Output: alpha = 0.10
```

APPENDIX B: TERMINOLOGY

ARC Principle: Artificial Recursive Creation. The principle that capability scales super-linearly with recursive depth in systems employing sequential self-referential processing.

Recursive Depth (R): The number of self-referential processing iterations. In AI systems, this may be proxied by chain-of-thought token count or reasoning step count.

Scaling Exponent (α): The power to which recursive depth is raised in the relationship $U = I \times R^\alpha$. Values greater than 1 indicate super-linear scaling.

Sequential Recursion: Recursive processing where each step builds explicitly on previous steps, enabling error detection and correction.

Parallel Recursion: Recursive processing where multiple independent branches are generated and aggregated (e.g., majority voting).

APPENDIX C: ADDRESSING THE OPERATIONALISATION CRITIQUE

A rigorous assessment (see v2 compass artifact) correctly notes that $U = I \times R^2$ "cannot currently be operationalised to scientific standards" when applied to cosmological scales (measuring universal complexity, cosmic entropy, etc.).

This paper addresses that critique by limiting scope:

We apply the ARC Principle specifically to **AI systems** where all variables are operationally defined:

Variable	Operational Definition	Measurement
U	Benchmark accuracy	AIME 2024 score
I	Single-pass accuracy	Zero-shot performance
R	Recursive depth	Thinking tokens or sample count
α	Scaling exponent	Calculated from error rate reduction

This limited scope is scientifically appropriate. Just as ideal gas law ($PV = nRT$) applies within its regime but fails at extremes, the ARC Principle may apply to AI systems without implying universal validity.

Extensions to cosmology, biology, or consciousness remain speculative until independently operationalised and tested.

Paper version: 5.0

Date: 17 January 2026

Priority established: Infinite Architects, published January 2026