

# EXPERIMENTAL RESEARCH PROGRAMME

## Testing the Equation $U = I \times R^2$

*A Concrete Protocol for Empirical Validation*

From the Framework:

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

Michael Darius Eastwood  
January 2026

## Executive Summary

This document presents a concrete experimental programme to test the equation  $U = I \times R^2$ , which proposes that structural complexity ( $U$ ) scales with intelligence ( $I$ ) multiplied by the square of recursive depth ( $R$ ). The programme identifies two primary domains where this relationship can be empirically tested with existing technology and data:

1. **Artificial Intelligence Systems** — measuring how capability scales with recursive reasoning depth (test-time compute, chain-of-thought iterations, meta-learning layers)
2. **Quantum Error Correction** — measuring how logical qubit stability scales with recursive code nesting depth

The programme specifies operational definitions for each variable, measurement protocols, quantitative predictions with explicit falsification thresholds, and a timeline for experimental validation.

**Central Hypothesis:** Capability/complexity scales quadratically (not linearly or exponentially) with recursive self-reference depth. Specifically, if we measure the exponent  $\alpha$  in the relationship  $U \propto R^\alpha$ , we predict  $\alpha = 2.0 \pm 0.3$ .

# 1. The Core Scientific Question

The equation  $U = I \times R^2$  makes a specific, testable claim: that the structural complexity of a system scales with its base information capacity multiplied by the square of its recursive depth. This implies that recursion has a multiplicative, not additive, effect on complexity—and specifically a quadratic one.

## 1.1 What Would Validation Mean?

If the equation holds empirically across multiple domains, it would establish that:

- Recursive self-reference is a fundamental amplifier of complexity
- The amplification follows a specific mathematical form (quadratic, not linear or exponential)
- Intelligence, defined as compression efficiency or predictive capacity, is the base unit that recursion amplifies

## 1.2 What Would Falsification Mean?

The equation would be falsified if:

- The measured exponent  $\alpha$  is consistently outside the range [1.7, 2.3]
- Different domains show radically different exponents with no unifying pattern
- The relationship is not multiplicative ( $I$  and  $R$  contribute additively rather than multiplicatively)

## 2. Domain 1: Artificial Intelligence Systems

AI systems provide the most tractable domain for testing the equation because all variables can be precisely controlled and measured.

### 2.1 Operational Definitions

Variable	Operational Definition	Measurement Protocol
U (Capability)	Benchmark performance score normalised across task difficulty	MMLU, GSM8K, MATH, ARC, HumanEval scores aggregated with difficulty weighting
I (Base Intelligence)	Single-pass prediction accuracy without chain-of-thought	Zero-shot performance on benchmarks with no reasoning time
R (Recursive Depth)	Number of self-referential reasoning iterations	Count of: (a) chain-of-thought steps, (b) self-correction loops, (c) meta-cognitive layers

### 2.2 Experimental Protocol: Test-Time Compute Scaling

#### 2.2.1 Rationale

OpenAI's o1 model demonstrates that "thinking longer" improves performance. The question is whether this improvement scales linearly, quadratically, or exponentially with thinking depth. This directly tests the  $R^2$  hypothesis.

#### 2.2.2 Experimental Design

**Independent Variable:** Number of reasoning iterations (R), controlled via:

- Chain-of-thought token budget (100, 500, 1000, 2000, 4000, 8000 tokens)
- Number of self-revision cycles (1, 2, 3, 4, 5)

**Dependent Variable:** Benchmark accuracy (U), measured across difficulty levels.

**Control:** Base model capability (I), held constant by using the same model checkpoint.

#### 2.2.3 Prediction

If  $U = I \times R^2$  holds, then for fixed I:

$$U \propto R^\alpha \text{ where } \alpha = 2.0 \pm 0.3$$

#### 2.2.4 Falsification Threshold

Measured Exponent $\alpha$	Interpretation	Status
$\alpha < 1.5$	Sub-linear scaling; recursion has diminishing returns	FALSIFIED
$1.5 \leq \alpha < 1.7$	Weakly super-linear; possible systematic error	WEAK SUPPORT
$1.7 \leq \alpha \leq 2.3$	Quadratic scaling confirmed	SUPPORTED
$2.3 < \alpha < 3.0$	Super-quadratic; cubic or higher relationship	PARTIAL SUPPORT
$\alpha \geq 3.0$	Exponential scaling; fundamentally different relationship	FALSIFIED

## 2.3 Existing Data to Analyse

Substantial data already exists that can be reanalysed through this framework:

3. **Kaplan et al. (2020) scaling laws:**  $\text{Loss} \propto N^{-0.076} \times D^{-0.095}$  — can we decompose  $N$  into  $I \times R^2$  components?
4. **OpenAI o1 technical report:** Performance scales with "time spent thinking" — extract the exponent
5. **DeepSeek-R1 ablations:** Chain-of-thought length vs. accuracy curves
6. **Encode-Think-Decode (ETD) paper:** +28.4% accuracy on GSM8K, +36% on MATH with recursive layer iteration — extract scaling curves

### 3. Domain 2: Quantum Error Correction

Quantum error correction provides a second, independent domain with precise numerical data already available.

#### 3.1 Operational Definitions

Variable	Operational Definition	Measurement Protocol
U (Stability)	Logical qubit lifetime / coherence time	Time until logical error probability exceeds threshold
I (Base Quality)	Physical qubit fidelity	Single physical qubit coherence time and gate fidelity
R (Recursive Depth)	Error correction code distance	Grid size: d=3 (3×3), d=5 (5×5), d=7 (7×7), etc.

#### 3.2 Existing Data: Google Willow Results

Google's Willow chip (December 2024) provides precise data on error suppression scaling:

Code Distance	Physical Qubits	Error Rate (per cycle)	Suppression Factor $\Lambda$
d = 3	17	~0.6%	—
d = 5	49	~0.3%	2.0×
d = 7	101	0.143%	2.14×

#### 3.3 Analysis Through the $U = I \times R^2$ Framework

##### 3.3.1 The Observed Relationship

Willow shows that each increase in code distance by 2 reduces the error rate by a factor of ~2.14. This means:

$$\epsilon_{\text{logical}} \propto \Lambda^{-d/2} \propto (1/2.14)^{(d-3)/2}$$

Converting to our framework where stability  $S = 1/\epsilon$ :

$$S \propto \Lambda^{d/2} \propto 2.14^{(d-3)/2}$$

##### 3.3.2 Testing the Quadratic Hypothesis

If  $U = I \times R^2$  holds with  $R = d$  (code distance), we would expect:

$$S \propto d^2$$

But the observed relationship is exponential in  $d$ , not quadratic. This requires interpretation:

- **Option A:** The recursion variable  $R$  is not the code distance  $d$  directly, but rather  $\log(d)$ . Then  $R^2 = (\log d)^2$  and the relationship becomes polynomial in  $\log d$ .
- **Option B:** Quantum error correction represents a different scaling regime where the squaring happens at each level of concatenation, producing  $\Lambda^{d^2}$  rather than  $d^2$ .
- **Option C:** The equation does not hold in the quantum domain — this would be partial falsification.

### 3.3.3 *Critical Test for Concatenated Codes*

Concatenated quantum codes provide a cleaner test. For  $L$  levels of concatenation:

$$\varepsilon_L \propto \varepsilon_0^{2^L} \text{ (error suppression is doubly exponential)}$$

This is dramatically stronger than quadratic. If  $U = I \times R^2$  holds with  $R = L$  (nesting depth), we would expect:

$$S \propto L^2 \text{ (stability scales quadratically with nesting)}$$

The prediction is therefore:  $\log(\log(1/\varepsilon))$  should scale linearly with  $L$ , while  $\log(S)$  should scale quadratically with  $L$ . These make different predictions that can be empirically distinguished.

## 4. Summary of Quantitative Predictions

ID	Prediction	Mathematical Form	Falsification Threshold
P1	AI capability scales quadratically with chain-of-thought depth	Accuracy $\propto (\text{CoT steps})^\alpha$ , $\alpha \in [1.7, 2.3]$	$\alpha < 1.5$ or $\alpha > 2.5$
P2	Meta-learning efficiency scales with meta-learning depth squared	Few-shot acc. $\propto (\text{meta-layers})^\alpha$	$\alpha < 1.5$ or $\alpha > 2.5$
P3	Recurrent depth models show quadratic improvement per iteration	Performance gain $\propto \text{iterations}^\alpha$	$\alpha < 1.5$ or $\alpha > 2.5$
P4	QEC concatenation shows $R^2$ in log-stability	$\log(\text{stability}) \propto L^\alpha$	$\alpha \approx 1$ (linear) or doubly exponential
P5	Cross-domain consistency: exponent $\alpha$ is similar across AI and QEC	$ \alpha_{\text{AI}} - \alpha_{\text{QEC}}  < 0.5$	$ \alpha_{\text{AI}} - \alpha_{\text{QEC}}  > 1.0$

## 5. Research Timeline

### Phase 1: Data Collection and Reanalysis (Months 1-3)

- Compile existing scaling law data from published papers
- Extract chain-of-thought scaling curves from o1, DeepSeek-R1, QwQ reports
- Reanalyse Willow data through  $U = I \times R^2$  framework

### Phase 2: Controlled Experiments (Months 4-9)

- Run controlled ablations varying only recursive depth while holding model constant
- Test across multiple benchmarks to check consistency
- Measure exponent  $\alpha$  with confidence intervals

### **Phase 3: Publication and Peer Review (Months 10-12)**

- Write up results with full methodology and data
- Submit to arXiv for community review
- Submit to peer-reviewed venue (NeurIPS, Nature Machine Intelligence, or Physical Review X)

## 6. Conclusion: From Speculation to Science

This experimental programme transforms  $U = I \times R^2$  from a philosophical claim into a testable hypothesis. The key innovations are:

7. **Operational Definitions:** Each variable has a concrete measurement protocol that yields numerical values.
8. **Quantitative Predictions:** The exponent  $\alpha$  should equal  $2.0 \pm 0.3$ , not merely "quadratic" but a specific numerical range.
9. **Explicit Falsification:** If  $\alpha < 1.5$  or  $\alpha > 2.5$  consistently, the hypothesis is falsified.
10. **Cross-Domain Testing:** The same relationship should hold across AI and quantum systems, providing independent validation.

The equation may prove correct, partially correct, or entirely wrong. But unlike metaphysical speculation, this programme will yield a definite answer. That is what makes it science.

*"The equation may prove wrong, trivial, or profound — but only after this transformation can science answer the question."*

## Appendix: Key References

Scaling Laws:

- Kaplan et al. (2020). Scaling Laws for Neural Language Models.  
arXiv:2001.08361
- Hoffmann et al. (2022). Training Compute-Optimal Large Language Models.  
arXiv:2203.15556

Test-Time Compute:

- OpenAI (2024). Learning to Reason with LLMs.  
[openai.com/index/learning-to-reason-with-langs](https://openai.com/index/learning-to-reason-with-langs)
- Geiping et al. (2025). Encode-Think-Decode: Scaling Test-Time Reasoning.  
arXiv:2510.07358

Quantum Error Correction:

- Google Quantum AI (2024). Quantum Error Correction Below the Surface Code Threshold. *Nature* 638, 920-926

Information Physics:

- Landauer (1961). Irreversibility and Heat Generation in the Computing Process.  
IBM J. Res. Dev. 5, 183-191
- Bérut et al. (2012). Experimental Verification of Landauer's Principle. Nature 483, 187-189