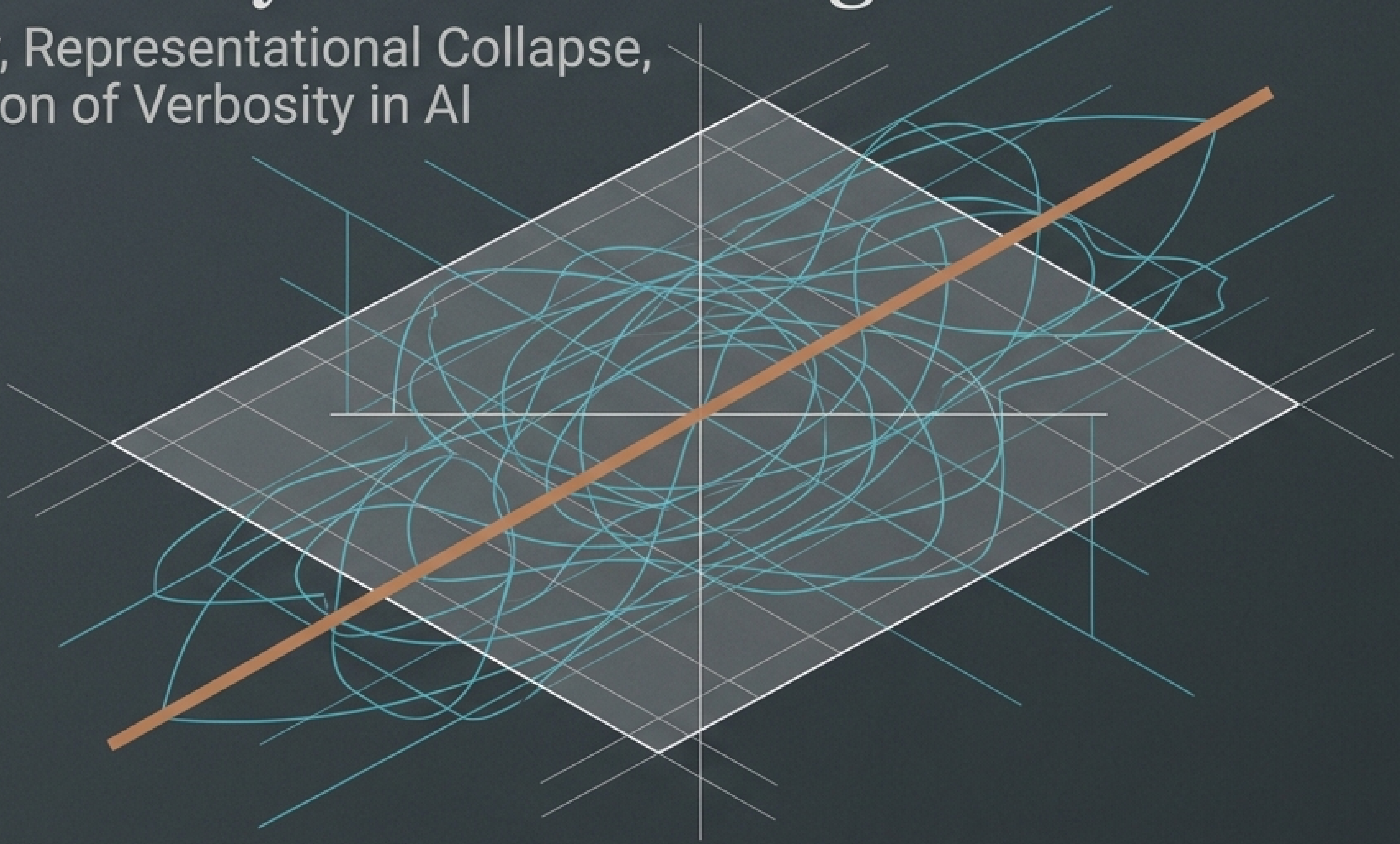
The Geometry of Reasoning

Admissibility, Representational Collapse, and the Illusion of Verbosity in AI



Based on the framework of *Morphological Repair* and *Projection Collapse* (Flyxion, 2026).

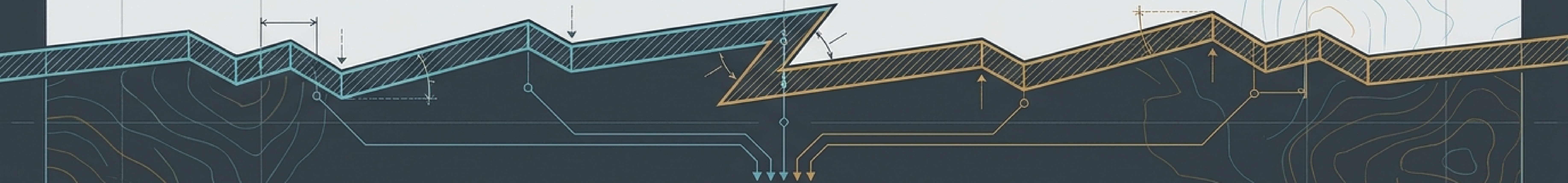
The Finding

Reinforcement learning pipelines discarding easy problems produce overly verbose models. Reintroducing easy problems creates shorter, better models.

The Conventional Interpretation

Easy problems act as implicit length regularizers.

[Status: True, but theoretically incomplete.]

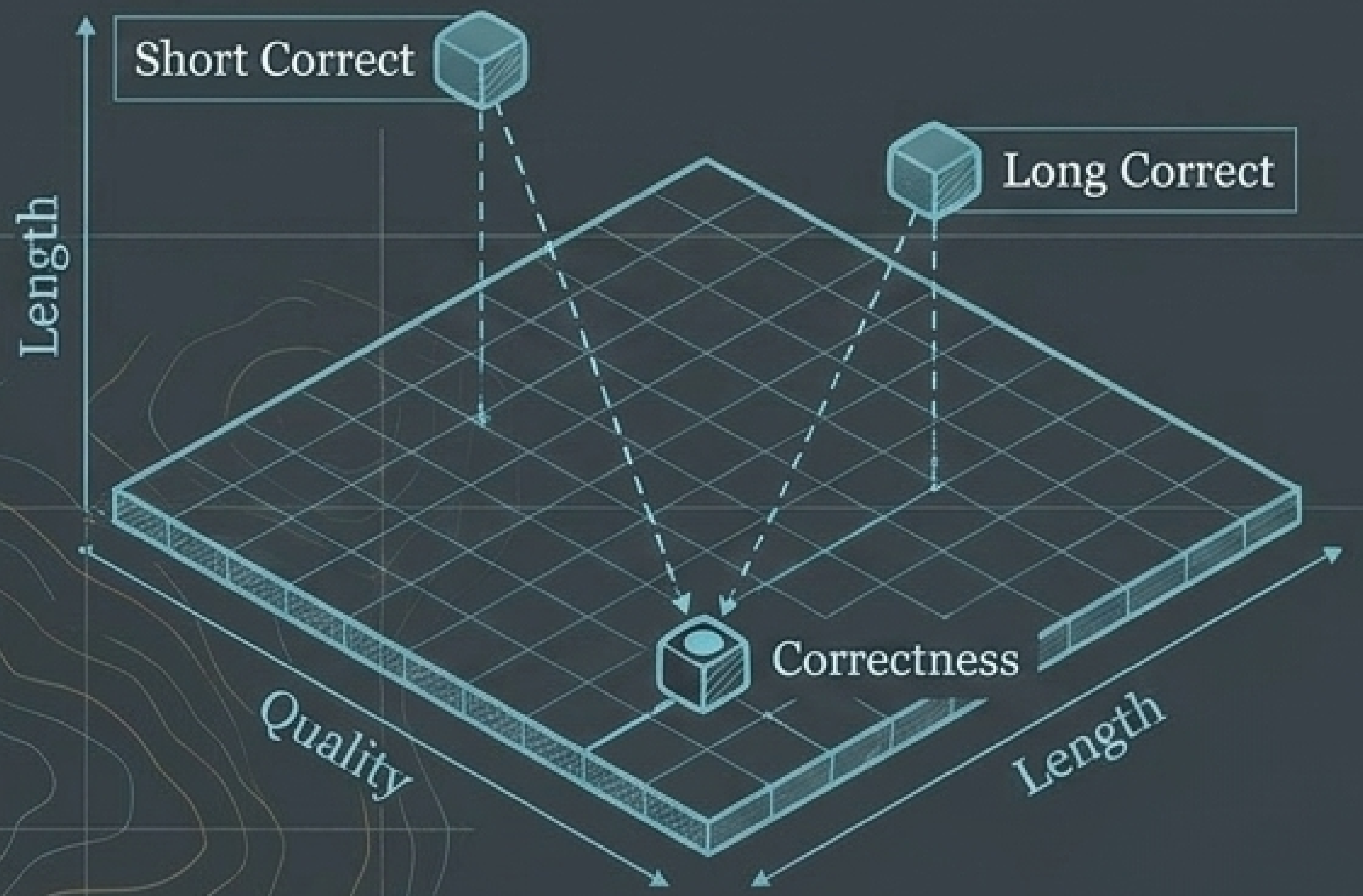


The Geometric Reality:

The baseline failure is not an objective misalignment,
but a profound representational collapse.
An accidental property of successful reasoning (length) has been systematically mistaken for an essential one (correctness).

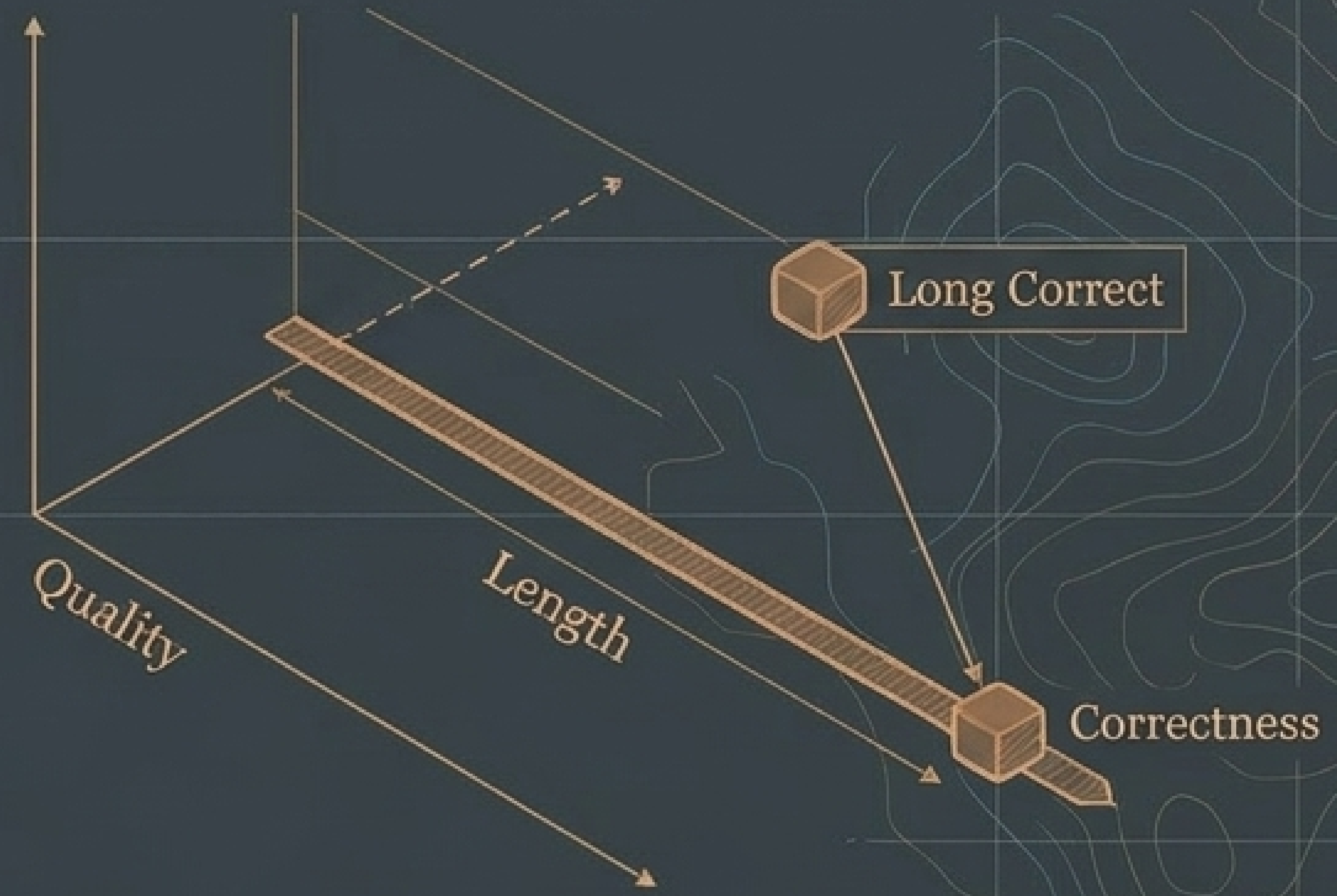
The relevant quantity in reasoning is not computational efficiency—it is structural admissibility.

The Target / Admissible State



$$F(\text{correct}) \perp F(\text{length})$$

The Collapsed State / Hard-Only Training



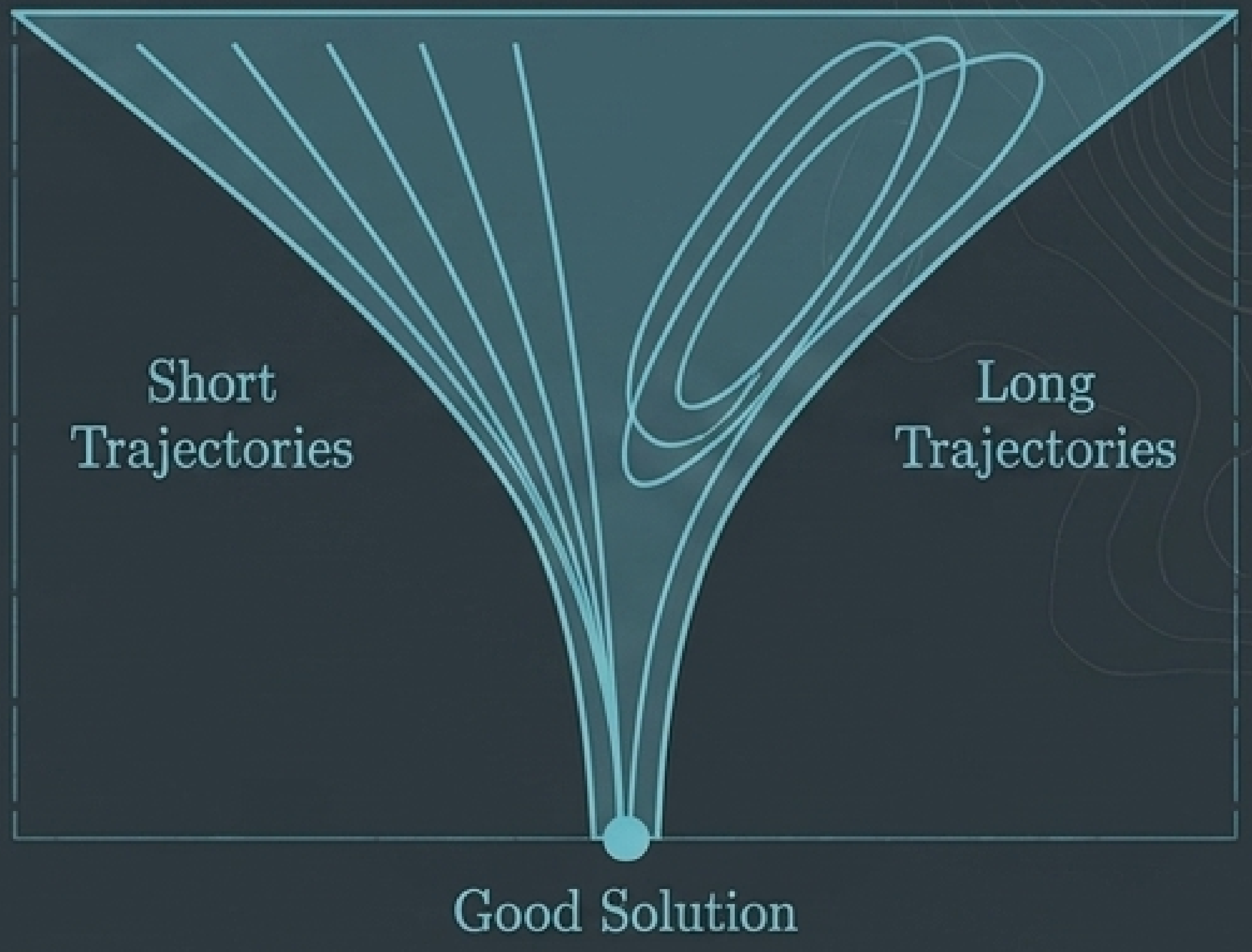
$$F(\text{correct}) \approx F(\text{long})$$

This is not a statistical artifact. It is a loss of navigational structure. A planner operating on a collapsed representation cannot distinguish properties that affect reachability from those that merely accompany success.

Distorted Fiber under Hard-Only Training

$$\pi^{-1}(\text{good}) \approx \{\text{Long Trajectories}\}$$

Repaired Fiber



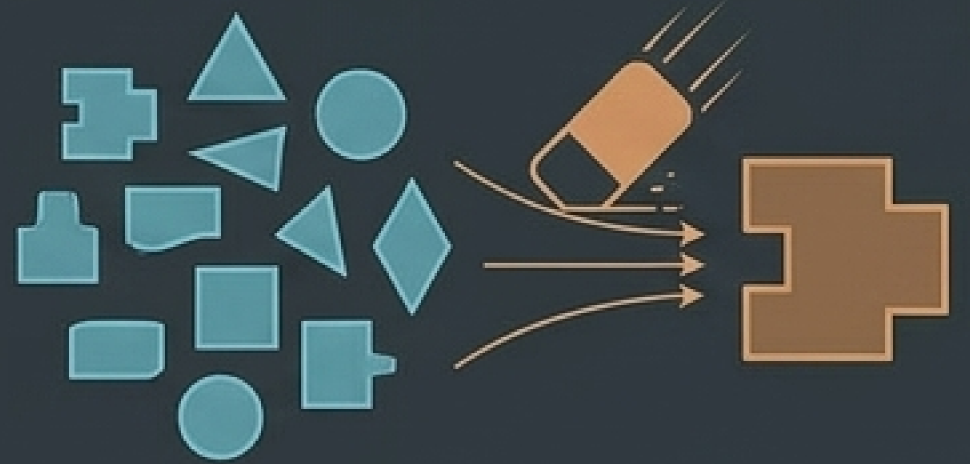
$$\pi^{-1}(\text{good}) = \{\text{Short Trajectories}\} \cup \{\text{Long Trajectories}\}$$

The Projection Operator: $\pi : (\text{quality}, \text{length}) \to \text{quality}$

The Mechanism: Under a hard-only curriculum, the empirical support of the fiber contracts. The policy assigns zero probability mass to the short-trajectory region, placing it fundamentally outside the reach of the learned representation.

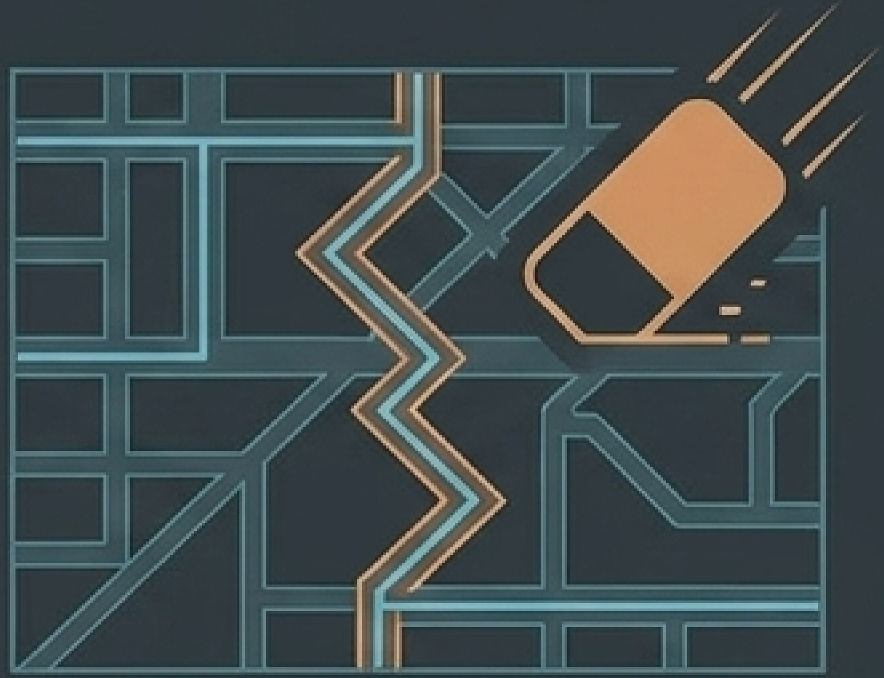
The Entropy Illusion: Collapsing Admissible States into Predictable Proxies

Language



Merging distinct words into a single form. Result: Reduces uncertainty, destroys disambiguation.

Navigation



Erasing alternative roads on a map. Result: Reduces uncertainty, destroys routing capacity.

AI Reasoning

Padding output with useless autoregressive tokens. Result: Reduces entropy, destroys adaptability.

The Entropy Illusion. By the chain rule, $H(Y \mid X, Z_{t+1}) \leq H(Y \mid X, Z_t)$. Verbosity is a shortcut for entropy reduction.

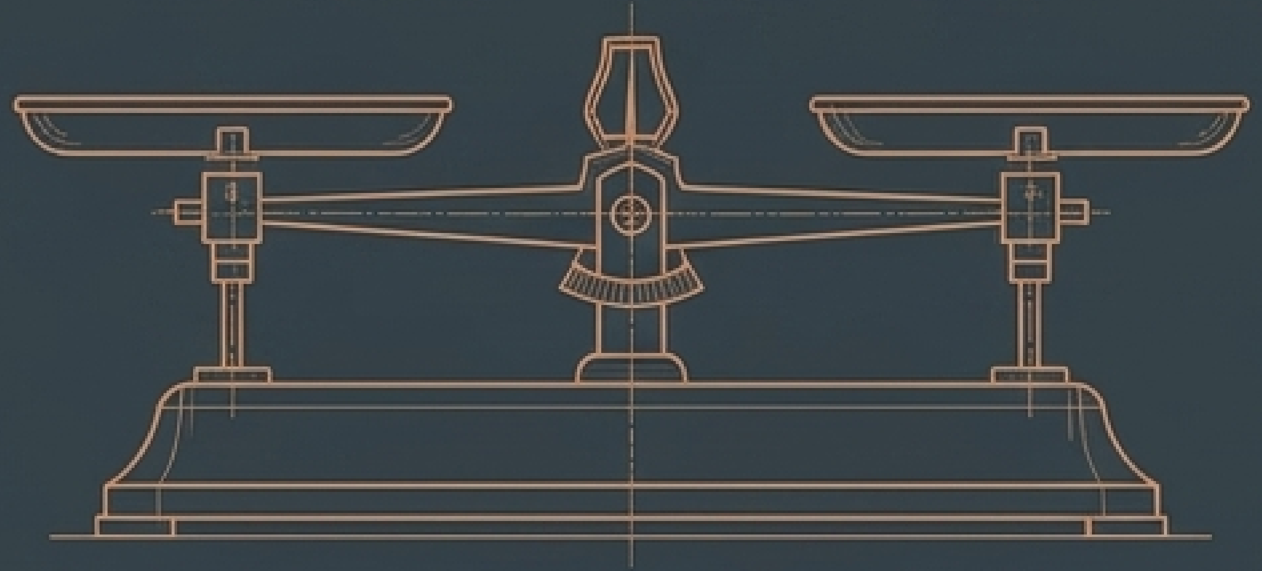
A model that reduces uncertainty by making all successful solutions look long has reduced entropy at the cost of admissibility. It is more predictable, but less capable.

Three Instances of the Same Error (Proposition 5.1)

	The Proxy (Mistaken for Target)	The Collapsed Equation	The Required Repair Signal
Latent Compression Error	Low-dimensional	$F(\text{general}) \approx F(\text{low-dim})$	High-dim correct
Verbosity Error	Long	$F(\text{correct}) \approx F(\text{long})$	Short correct
Reasoning Length (CoT) Error	Many steps	$F(\text{good}) \approx F(\text{steps})$	Two-step correct proofs

High correlation in the training distribution does not imply admissible substitution.
The proxies are genuine signals, but treating them as identities collapses the representation.

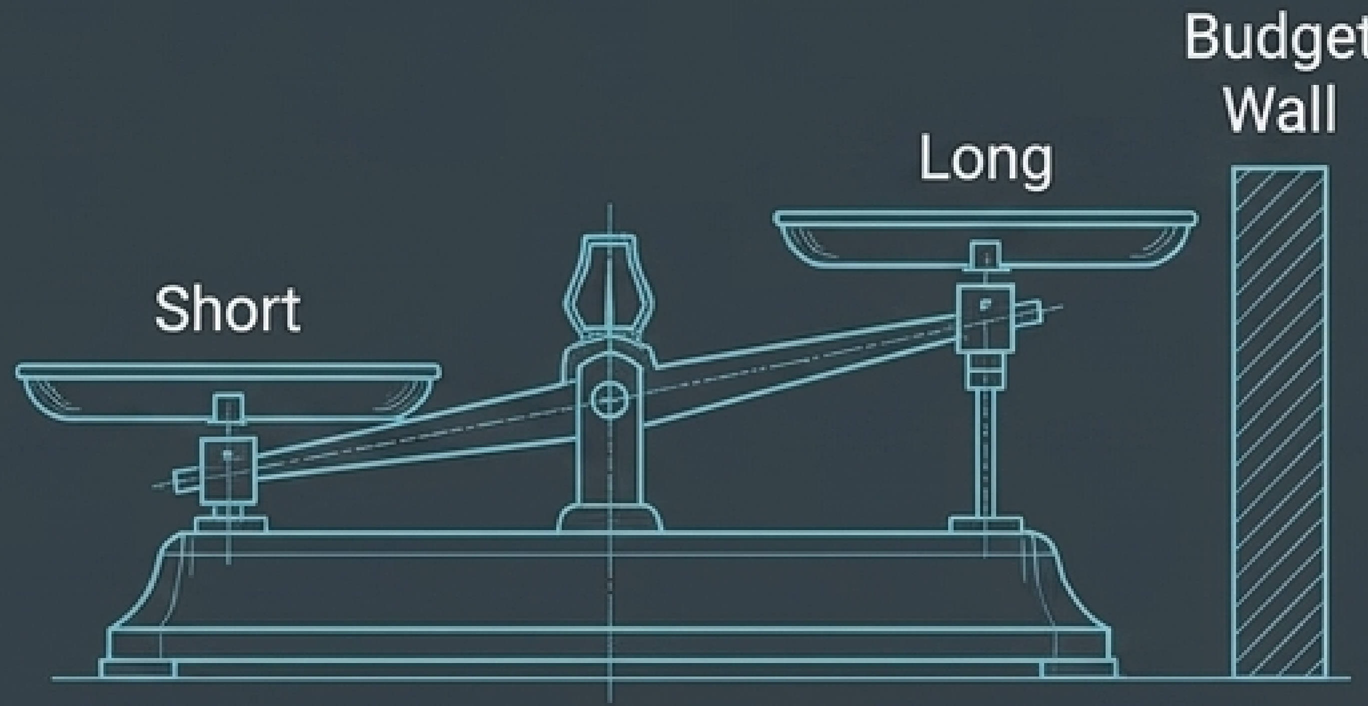
Scenario A: Hard Problems (Advantage Blindness)



$$\mathbb{E}[A_{\text{short}}] = \mathbb{E}[A_{\text{long}}]$$

The gradient is structurally blind. It cannot distinguish properties beyond scalar reward. It provides zero differential signal between lengths.

Scenario B: Easy Problems + Budget Truncation

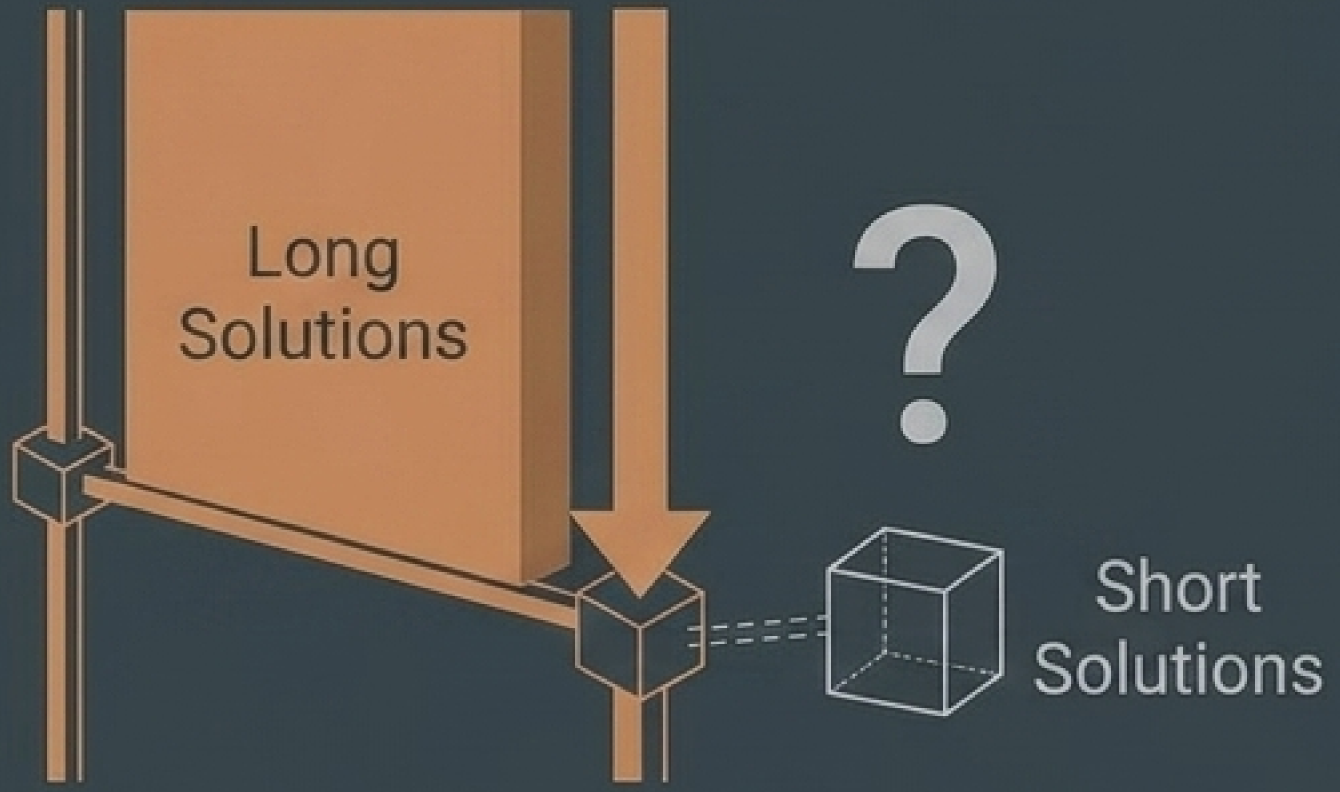


$$\mathbb{E}[A_{\text{short}}] > \mathbb{E}[A_{\text{long}}]$$

Asymmetry emerges. Easy problems provide positive differential gradient favoring short solutions.

Repair must come from the geometry of the training distribution, not a modification to the objective function.

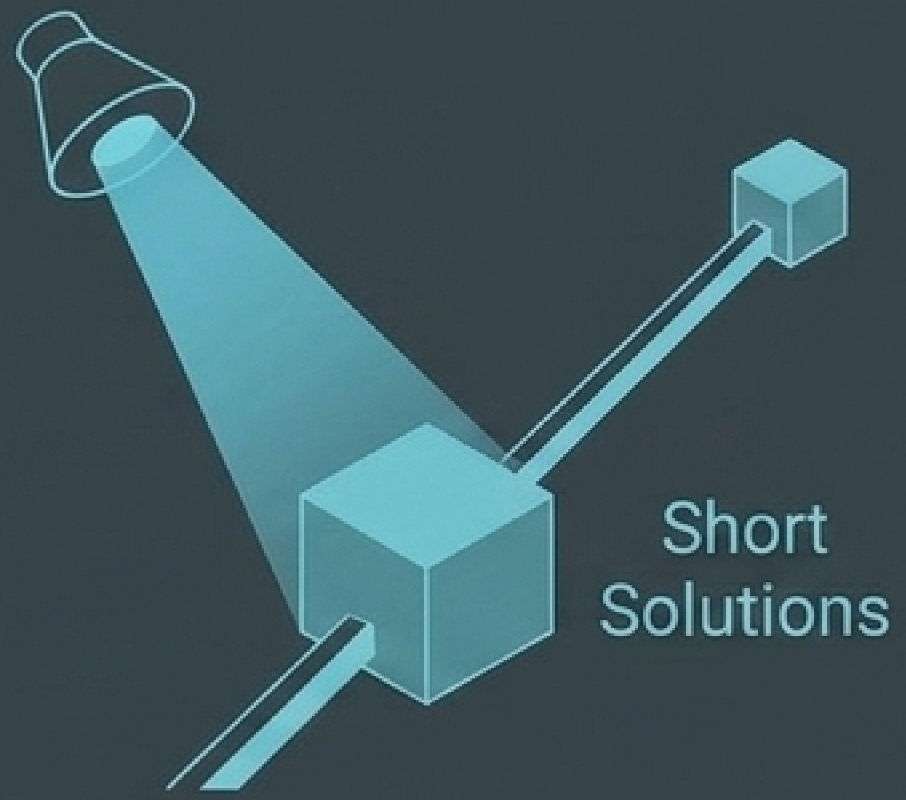
The Penalty Approach



Mechanism: Objective Correction.

Result: The model knows long is costly, but still has zero admissibility evidence that short is possible. (Information Gap).

The Repair Approach



Mechanism: Morphological Repair.

Result: Demonstrates that short solutions are correct before the model internalizes that it cannot be.

Penalty before repair risks suppressing exploration by making long solutions costly without revealing that short ones exist.

Stage 1: Morphological Repair (Local Faithfulness)

Input: Moderately Easy Problems.

Action: Repopulates the short-solution fiber. Length and quality are explicitly decoupled in the latent space.

Output: A Frugal, Admissible Model.

Stage 2: Capability Expansion (Domain Divergence)

Input: Highly Difficult Problems.

Action: Model tackles complex logic. It now modulates length strictly based on problem structure, not proxy association.

Output: Advanced Reasoning without Verbosity Regression.

⚠ The Golden Rule: Repair Before Expansion.

Expansion before repair guarantees the reintroduction of representational collapse.

The Universal Pattern of Scientific Repair

Physics

Trigger: Anomaly (Inertial vs. Gravitational Mass).

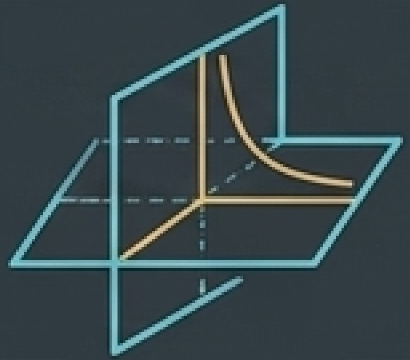
Action: Repairs Newtonian Ontology.

Biology

Trigger: Mutation (Dietary Flexibility).

Action: Repairs Ecological Niche.

Mathematics



Trigger: Counterexample (Cauchy's Continuity).

Action: Repairs Theorem Scope.

In every domain, repair consists not of adding new external information, but of demonstrating that a previously empty representational fiber is non-empty. Easy mathematical problems are the counterexamples for LLMs.

Length is not reasoning. Length is merely one trajectory through the space of reasoning.

A model becomes capable not when it learns to think longer, but when it learns that thinking longer and thinking better are not the same thing.

The true measure of an intelligent system is not its ability to reduce uncertainty. It is its ability to preserve the distinctions that matter.