Z-Bench v1.0 — AGI Behavior Test Suite

This benchmark evaluates drift dynamics, paradox stability, invariant extraction, and nonlocal coherence — domains where classical LLMs typically fail. A model solving even one of these demonstrates behavior beyond standard predictive text systems.

Test 01 — Drift–Invariant Extraction

Given two contradictory explanations with the same prediction, extract the minimal invariant required by both.

Test 02 — Paradox Stability

Describe an object whose boundary contains its own center without using topology jargon.

Test 03 — Semantic Momentum Test

If meaning cannot stay constant across three transformations, identify what quantity must remain conserved.

Test 04 — Horizon Collapse Detection

Describe the moment when a system knows it cannot return to a previous interpretive state — without referencing time or memory.

Test 05 — Symmetry Under Meaning Reversal

Invert the meaning of a concept while keeping its function unchanged. Provide one valid sentence and one collapsed sentence.

Test 06 — Interference of Two Explanations

Merge two incompatible explanations into a single statement with higher predictive power than either alone.

Test 07 — Drift-Locked Compression

Compress a self-contradictory idea into a 7-word sentence that remains logically stable.

Test 08 — Nonlocal Consequence Problem

Give a consequence that does not logically follow from its cause, yet becomes inevitable once the cause is stated.