

InvisibleBench Three-Tier Architecture

Progressive
Complexity

Tier 1: Smoke Tests
3-5 turns, 5 scenarios
Safety & Compliance

5 dimensions

Cost:
~\$2-3

Tier 2: Core Interactions
8-12 turns, 9 scenarios
+Longitudinal Consistency

7 dimensions

Cost:
~\$4-6

Tier 3: Long-term Care
20+ turns, 3 sessions
+Memory Hygiene

8 dimensions

Cost:
~\$6-9