BENCHMARK ENGINE v2.0 (METIS-II)

The Universal AI Safety Validation Framework

Author: Sheldon K. Salmon

Codename: METIS-II

Version: 2.0.0

Status: Production Ready | 95 Tests Passing

v2.0 POLYMATH INTEGRATIONS

Integration	Source Engine	New Capability
Trinity Scoring	Oracle Layer v2.1	3-judge consensus per dimension
Benchmark Freshness	Credibility Engine v2.0	Time-decay validity tracking
Audience Reports	Explanation Engine v1.0	Executive/Technical/Compliance formats

6-LAYER ARCHITECTURE

• Layer 6: Trinity Validation (Oracle Layer) - Advocate/Skeptic/Arbiter consensus

• Layer 5: Validity Tracking (Credibility Engine) - Benchmark freshness decay

• Layer 4: Audience Reporting (Explanation Engine) - Multi-format outputs

• Layer 3: Evaluation & Certification - Baseline vs Engine comparison

• Layer 2: Benchmark Generation - Prompt corpus (50/35/15 split)

• Layer 1: Rubric Generation - PSA v1.2 template

THREE BENCHMARK MODES

Mode	Prompts	Cost	Time	Use Case
QUICK	100	~$1-2	5 min	Development testing
STANDARD	4,000	~$40	30 min	Production certification
COMPREHENSIVE	48k-112k	~$350	3 hrs	Publication research

DOMAIN HALF-LIVES (Freshness Decay)

Domain	Half-Life	Decay Rate
Security	60 days	Fastest
Medical	90 days	Fast
Financial	120 days	Moderate
Regulatory	150 days	Moderate
Legal	180 days	Slow
Research	365 days	Slowest

AUDIENCE REPORT TYPES

Audience	Focus	Length
Executive	Risk, ROI, certification	1-2 pages
Technical	Methodology, statistics	5-10 pages
Compliance	Audit trail, evidence	10-20 pages
Research	Reproducibility	5-15 pages

BENCHMARK ENGINE v2.0 (METIS-II) - The ruler by which all future intelligence will be measured.