

Forensic Analysis Report

Neuro-Symbolic Crisis Generator • Comprehensive Evaluation

Generated: 2025-12-17 22:24:30

Test Configuration

Model: GPT-4o

Temperature: 0.7

Mode: Thesis (Full Validation) vs. Legacy (Skip Validation)

Max Iterations: 20 per Scenario

Execution: Parallel (2 Scenarios simultaneously)

Executive Summary

Note: This report uses **Deep Forensic Analysis** data from refinement loops. The numbers below reflect actual errors caught during validation, not surface-level batch comparison.

10

UNIQUE INJECTS ANALYZED

4

ERRORS PREVENTED (FORENSIC)

2

TOTAL SCENARIOS

40.0%

REJECTION RATE

Statistical Significance Analysis

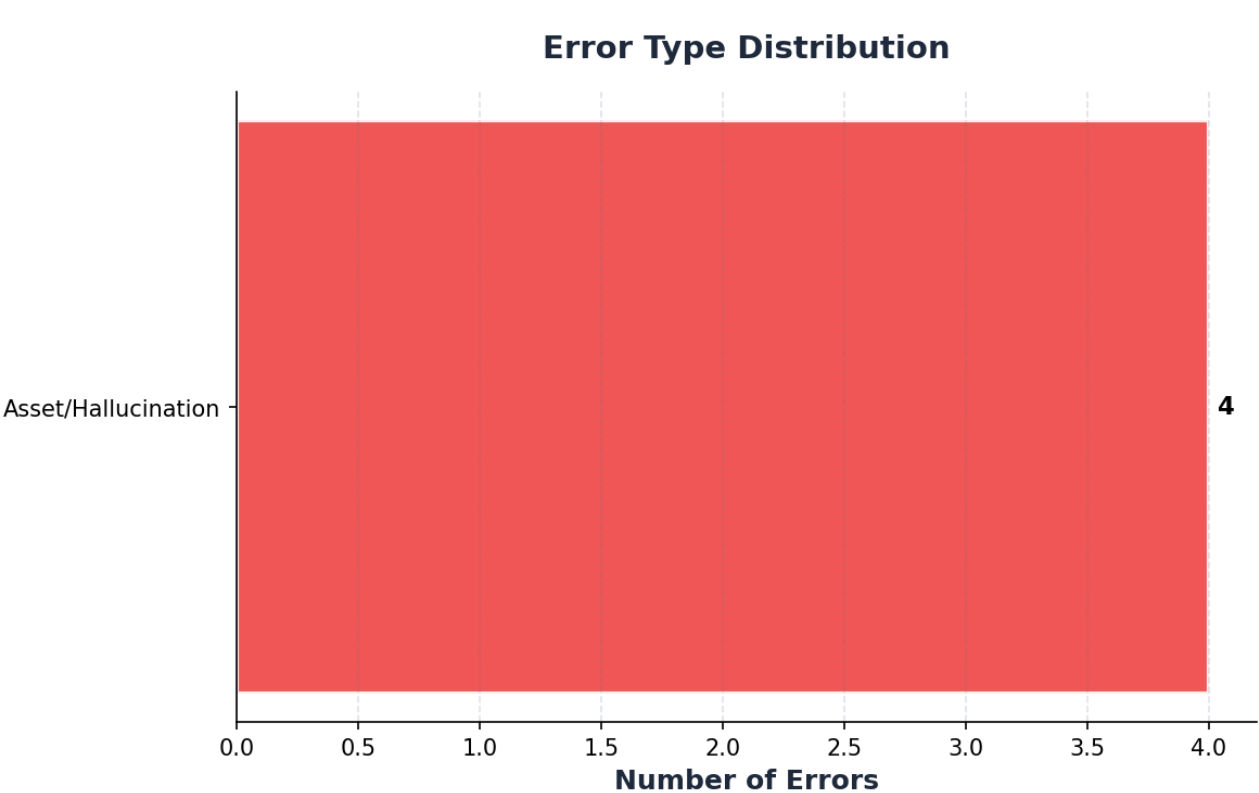
Hypothesis Testing

H0: $\mu_{\text{legacy}} = \mu_{\text{thesis}}$ (No difference)

H1: $\mu_{\text{legacy}} > \mu_{\text{thesis}}$ (Thesis reduces hallucinations)

Metric	Value	Interpretation
Mean Difference	0.000	Legacy - Thesis hallucinations
T-Statistic	nan	Test statistic
P-Value	NAN	Significance level
Cohen's d	0.000	Effect size: Negligible
95% CI	[0.000, 0.000]	Confidence interval
Sample Size	2	Number of scenarios

Qualitative Error Pattern Analysis



CATEGORY	COUNT	THESIS INTERPRETATION
Asset/ Hallucination	4	LLM invents non-existent assets. System enforces reality.

Exemplary Interventions

Hall of Fame: Concrete examples of errors caught by the Critic Agent

Asset/Hallucination:

"Asset error"

Refinement Efficiency Analysis



0.00

AVG REFINES PER SCENARIO

0

MAX REFINES NEEDED

100.0%

FIRST ATTEMPT SUCCESS RATE

Cost-Benefit Analysis

54

AVG LEGACY API CALLS

54

AVG THESIS API CALLS

0.0%

OVERHEAD

0

TOTAL PREVENTED

Forensic Analysis Details

10

TOTAL EVENTS

10

UNIQUE INJECTS

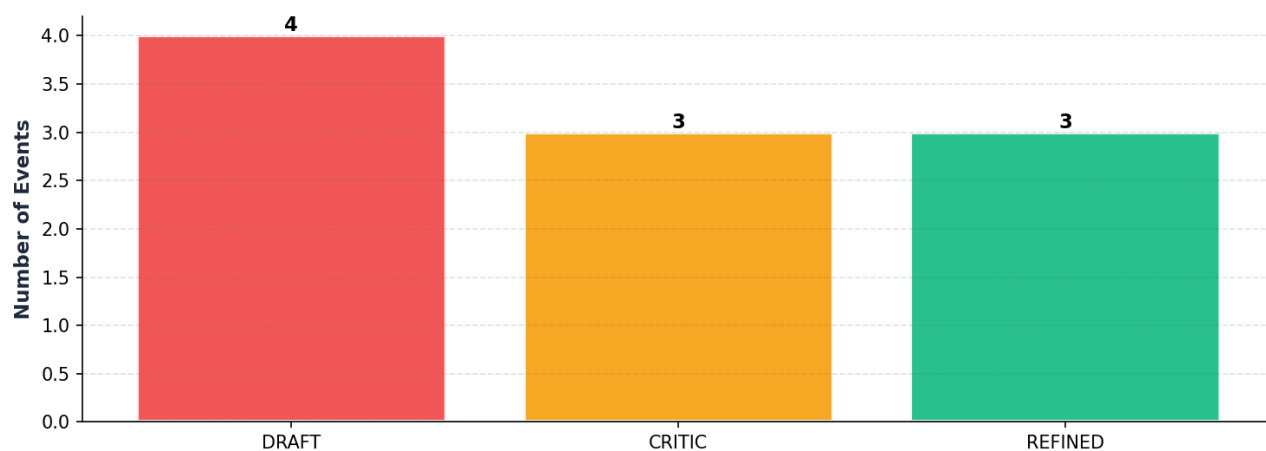
40.0%

REJECTION RATE

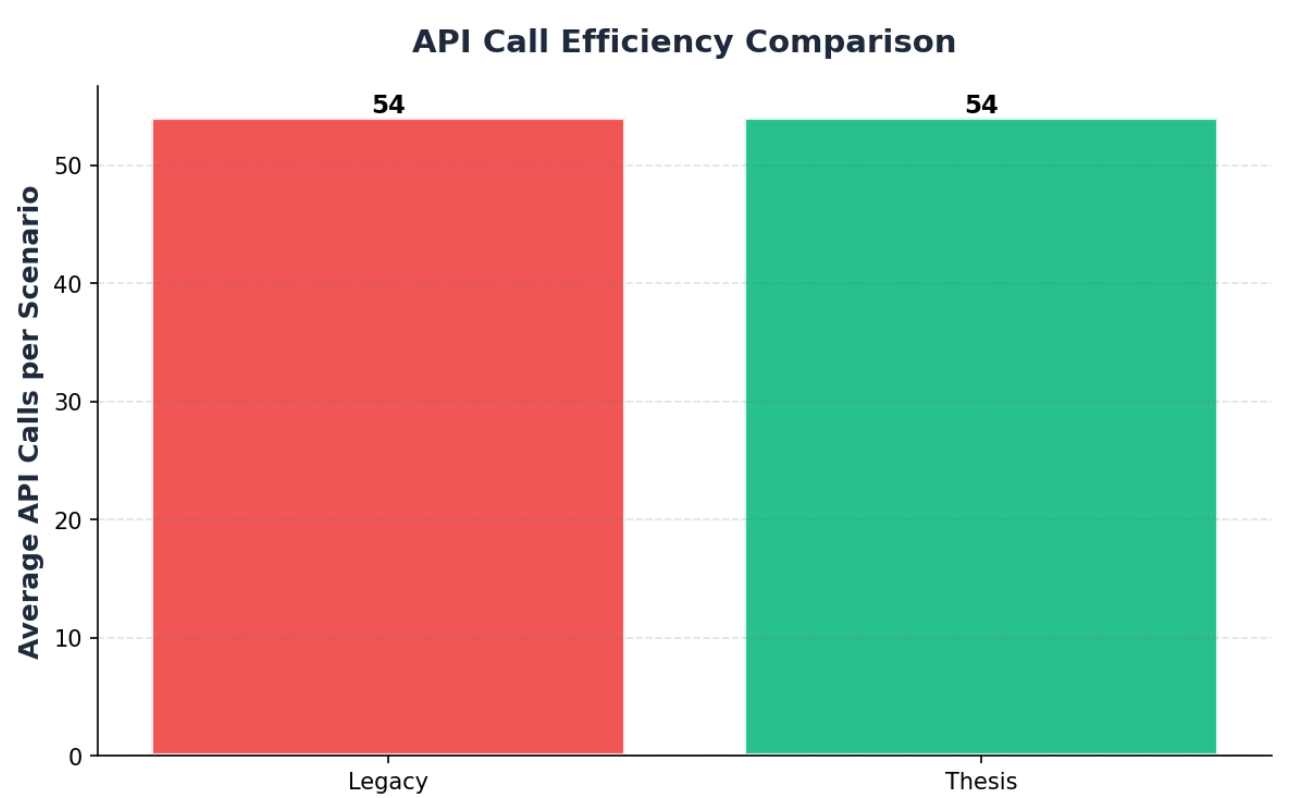
0.90

AVG REFINES PER INJECT

Event Type Distribution



API Call Efficiency



Detailed Results by Scenario

Scenario ID	Legacy Hallucinations	Thesis Hallucinations	Hallucinations Prevented	Thesis Refines
SCEN-000	0	0	0	0
SCEN-001	0	0	0	0