Forensic Analysis Report

Neuro-Symbolic Crisis Generator • Comprehensive Evaluation

Generated: 2025-12-17 22:51:54

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

10

UNIQUE INJECTS ANALYZED

4

ERRORS PREVENTED (FORENSIC)

2

TOTAL SCENARIOS

**40.0%**

REJECTION RATE

Statistical Significance Analysis

Hypothesis Testing
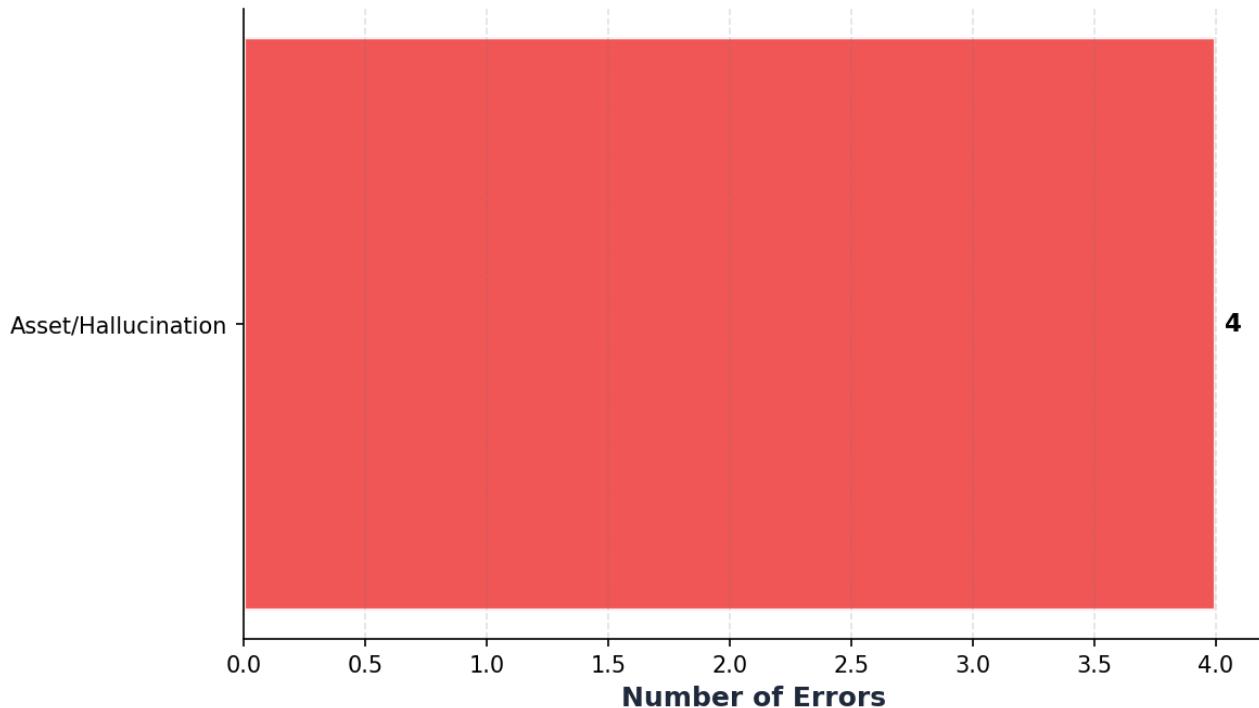
H0: μ_legacy = μ_thesis (No difference)

H1: μ_legacy > μ_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

Error Type Distribution

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

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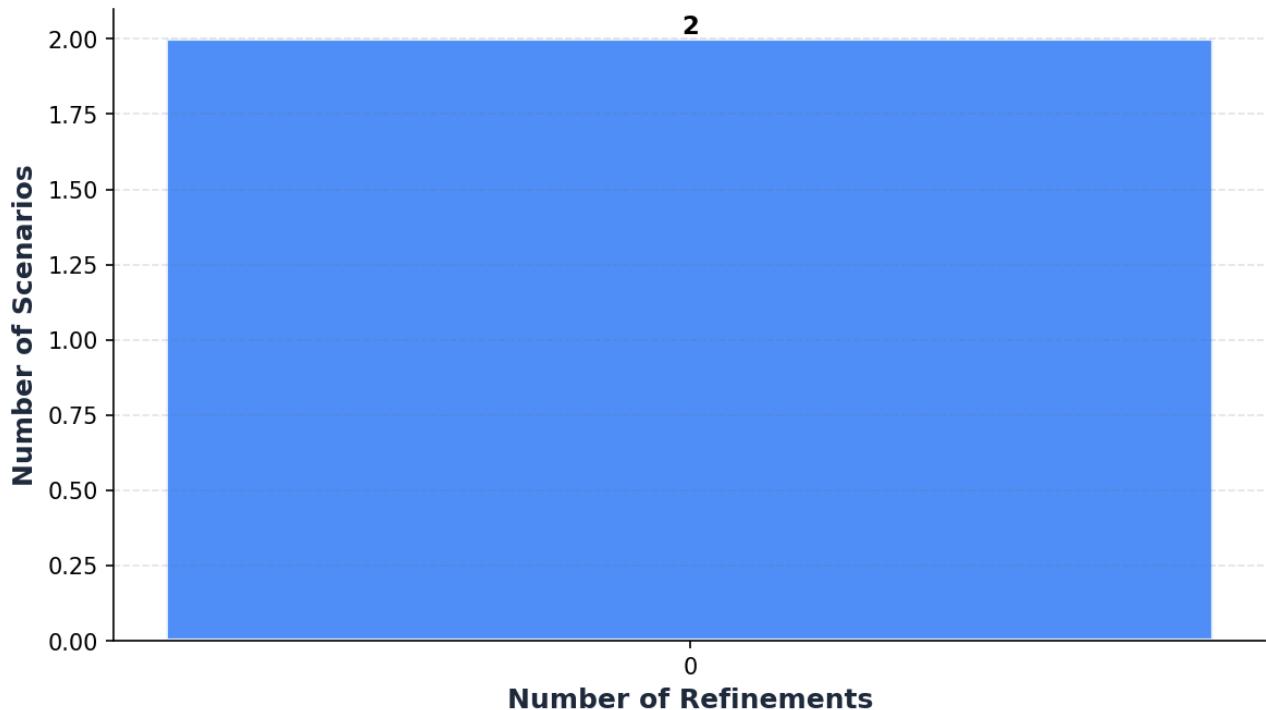
Hall of Fame: Concrete examples of errors caught by the Critic Agent

Asset/Hallucination:

"*Asset error*"

Refinement Efficiency Analysis

Refinement Distribution

Number of Scenarios (y-axis: 0.00 to 2.00)
Number of Refinements (x-axis: 0)

Bar value: 2

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
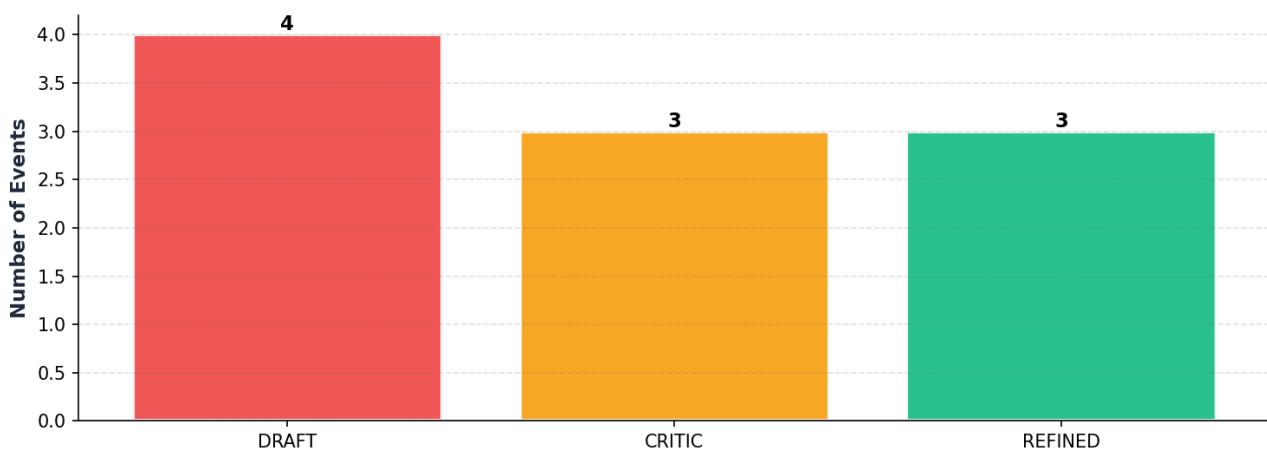
40.0%

REJECTION RATE

0.90
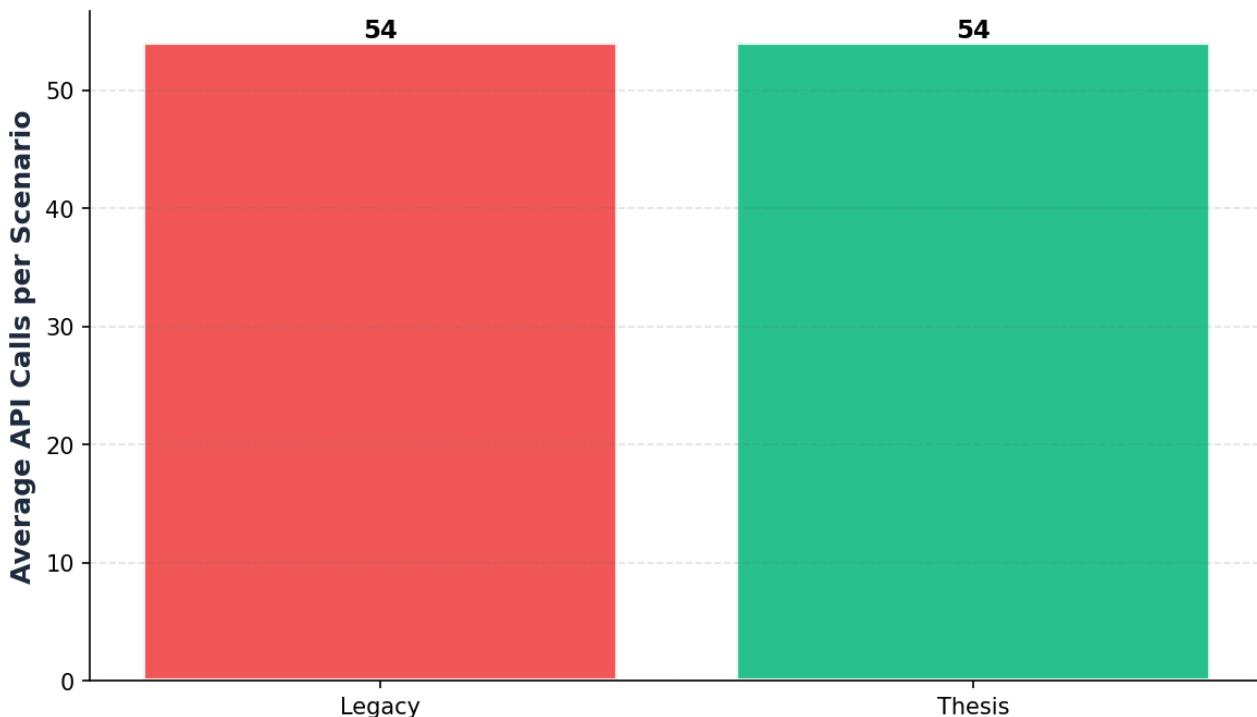
AVG REFINES PER INJECT

Event Type Distribution

API Call Efficiency

API Call Efficiency Comparison

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