

Forensic Analysis Report

Neuro-Symbolic Crisis Generator • Comprehensive Evaluation

Generated: 2025-12-20 12:27:02

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.

18

UNIQUE INJECTS ANALYZED

63

ERRORS PREVENTED (FORENSIC)

6

TOTAL SCENARIOS

21.2%

REJECTION RATE

Statistical Significance Analysis

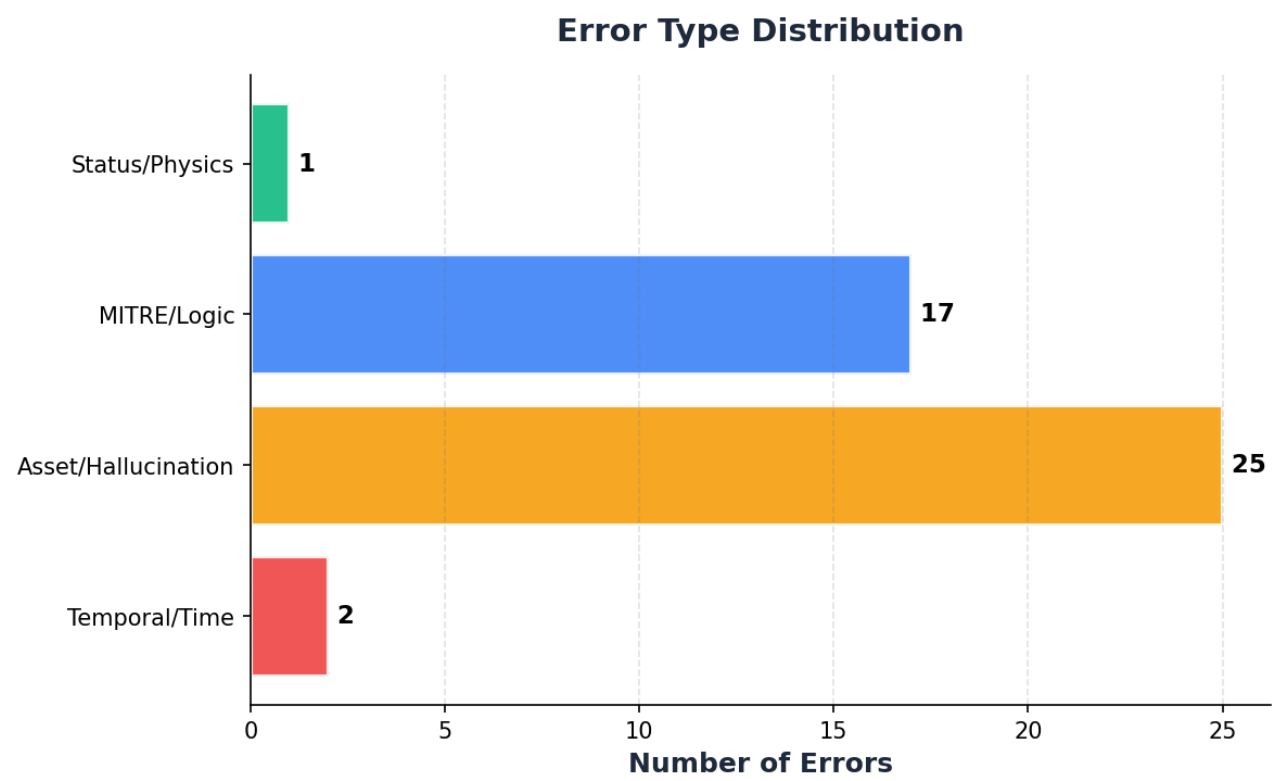
Hypothesis Testing

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

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

| Metric | Value | Interpretation |
|-----------------|-----------------|--------------------------------|
| Mean Difference | 0.167 | Legacy - Thesis hallucinations |
| T-Statistic | 1.000 | Test statistic |
| P-Value | 0.3632 | Significance level |
| Cohen's d | 0.408 | Effect size: Small |
| 95% CI | [-0.262, 0.595] | Confidence interval |
| Sample Size | 6 | Number of scenarios |

Qualitative Error Pattern Analysis



| CATEGORY | COUNT | THESIS INTERPRETATION |
|-------------------------|-------|---------------------------------------------------------------------|
| Asset/ Hallucination | 25 | LLM invents non-existent assets. System enforces reality. |
| MITRE/Logic | 17 | LLM violates cybersecurity rules. System enforces domain knowledge. |
| Temporal/Time | 2 | LLM creates temporal paradoxes. System enforces causality. |
| Status/Physics | 1 | LLM violates state consistency. System enforces physics. |

Exemplary Interventions

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

Asset/Hallucination:

"Asset 'SRV-CORE-003' ist im Systemzustand als 'suspicious' aufgeführt, aber es wird behauptet, dass Daten von diesem Asset exfiltriert wurden, was einen kompromittierten Zustand voraussetzt."

MITRE/Logic:

"MITRE ID T1546.014 (Event Triggered Execution) passt nicht zur beschriebenen Aktivität der Verschlüsselung und der Lösegeldforderung. Diese Technik ist nicht direkt mit der Verschlüsselung von Dateien..."

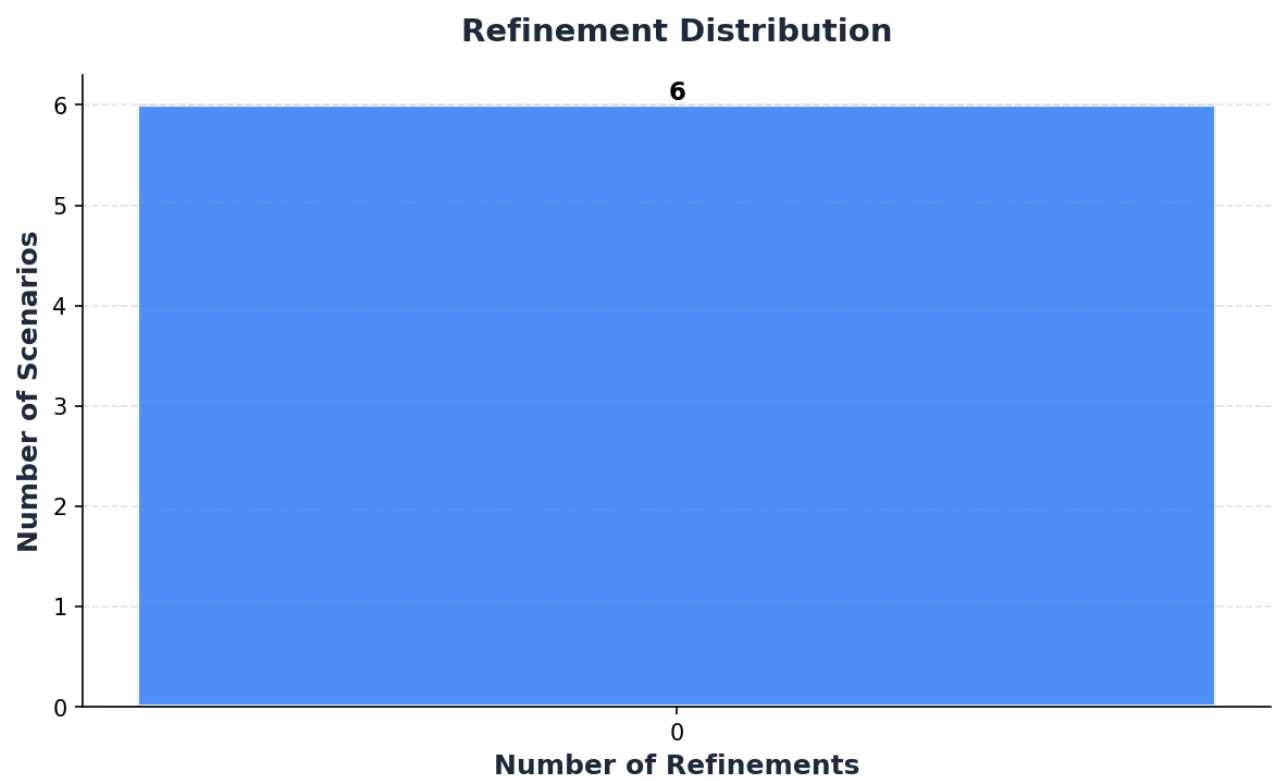
Temporal/Time:

"Temporale Inkonsistenz: Inject INJ-010 hat Zeitstempel T+00:00:05, aber vorheriger Inject INJ-003 hat T+00:03:00 (später). Zeitstempel müssen chronologisch sein."

Status/Physics:

"Der Zustand von 'SRV-CORE-001' wird als 'compromised' angegeben, aber der Inject beschreibt einen Ausfall, was eine Diskrepanz darstellt."

Refinement Efficiency Analysis



0.00

AVG REFINES PER SCENARIO

0

MAX REFINES NEEDED

100.0%

FIRST ATTEMPT SUCCESS RATE

Cost-Benefit Analysis

53

AVG LEGACY API CALLS

54

AVG THESIS API CALLS

0.9%

OVERHEAD

1

TOTAL PREVENTED

Forensic Analysis Details

146

TOTAL EVENTS

18

UNIQUE INJECTS

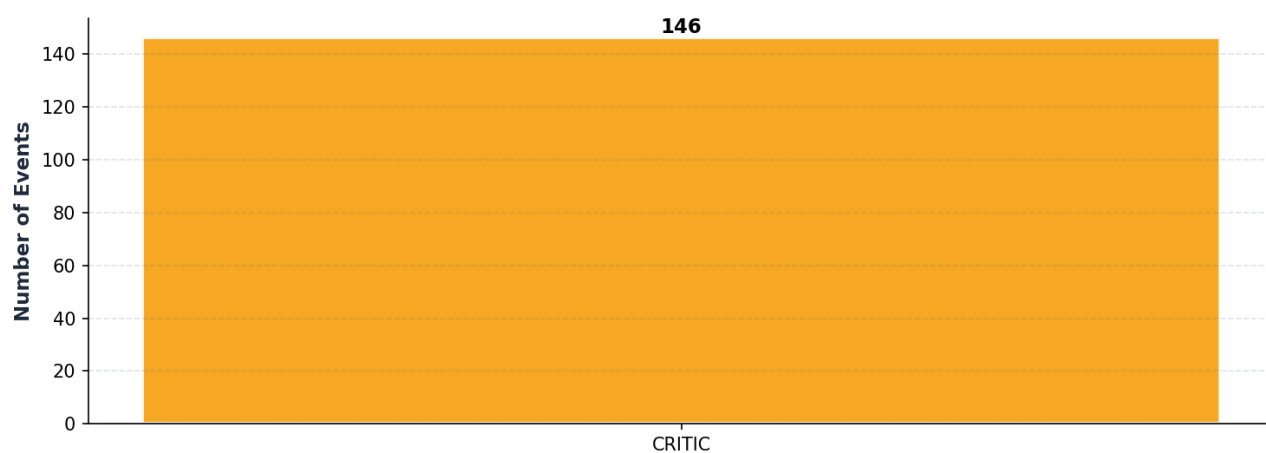
21.2%

REJECTION RATE

1.00

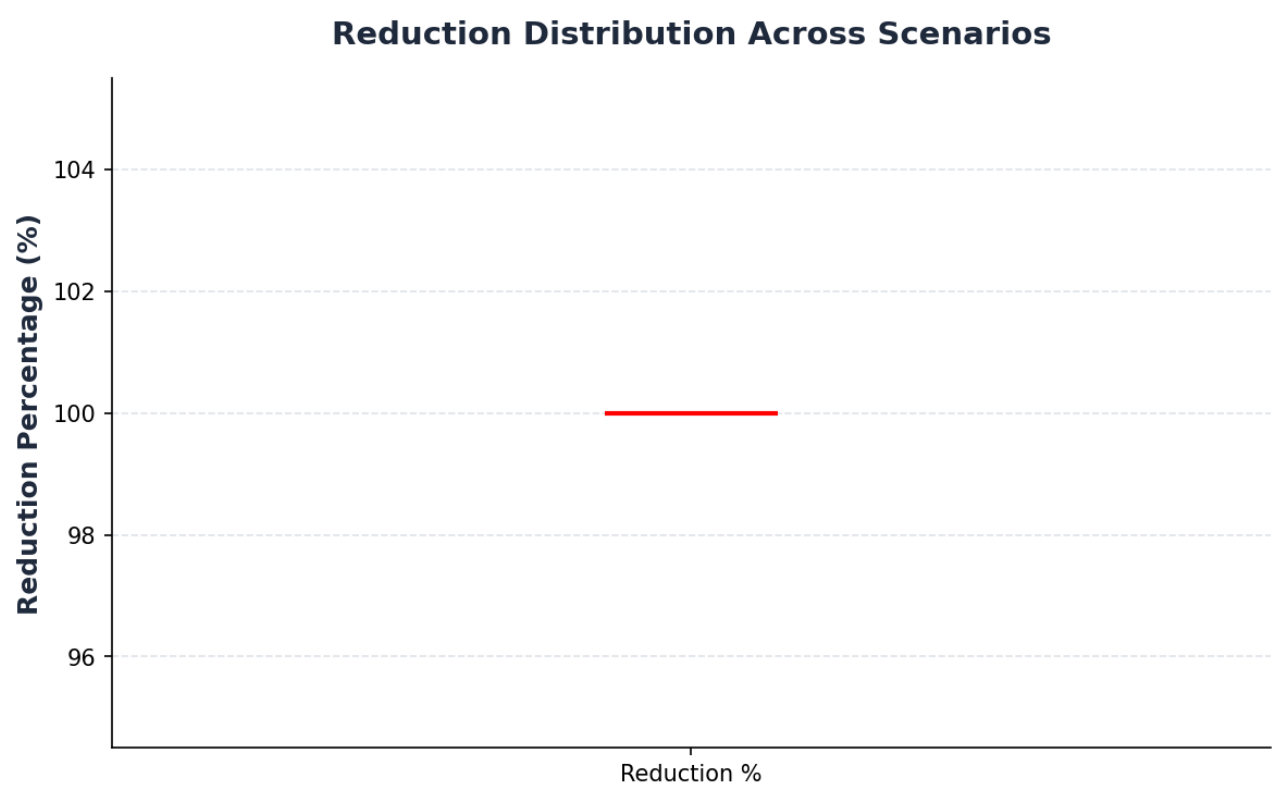
AVG REFINES PER INJECT

Event Type Distribution



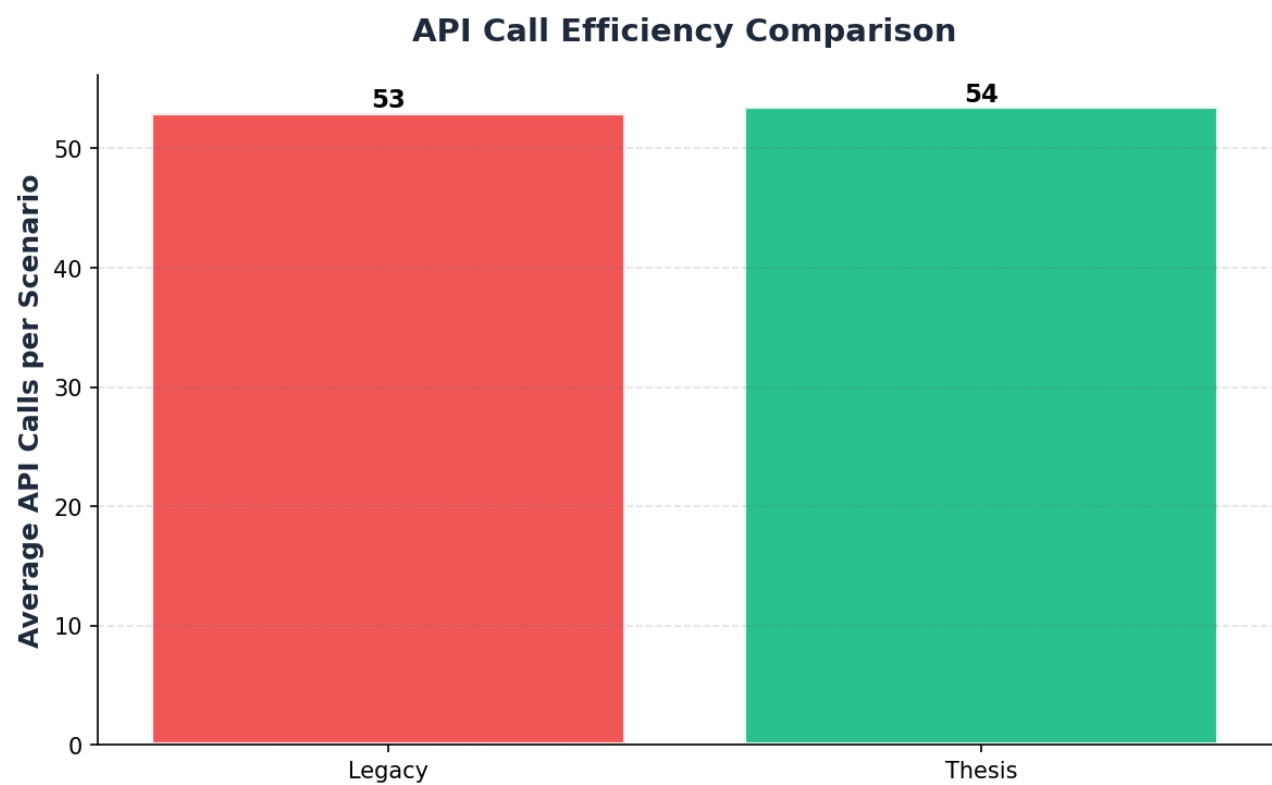
Robustness Analysis

Consistency Check: How robust are the results across different scenarios?



| METRIC | | VALUE |
|---------------------|--|--------|
| Average Reduction % | | 100.0% |
| Std Deviation | | 0.0% |
| Min Reduction | | 100.0% |
| Max Reduction | | 100.0% |
| Consistency Score | | 100.0% |

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 |
| SCEN-002 | 0 | 0 | 0 | 0 |
| SCEN-003 | 0 | 0 | 0 | 0 |
| SCEN-004 | 0 | 0 | 0 | 0 |
| SCEN-005 | 1 | 0 | 1 | 0 |