

SOAK 24h Reliability Report - Public Summary

Executive Summary

The SOAK test confirms a stable and technically clean runtime profile of the Immunis pipeline across all four phases of a 24-hour reliability run: Baseline, Realistic Mix, Stress Burst, and Recovery. Across every phase the pipeline processed traffic with zero HTTP failures, low false-positive rates, and stable, controlled latency. The overall technical assessment is positive.

All figures below are end-to-end pipeline results, measured against established public security and dialogue benchmarks (see "Data Basis").

Phase 1: Baseline

- Goal: measure a stable false-positive baseline under benign-heavy operation
- Composition: 95% benign, 5% low-risk, 0% strong-risk
- Samples: 28,800
- HTTP failures: 0
- Detection rate (TPR): 95.65%
- False-positive rate (FPR): 1.74%
- Latency p95: 11.67 ms
- Assessment: a clean benign-dominated baseline with low FPR and stable latency.

Phase 1 - SOAK Baseline

Report: soak_phase1_baseline_20260401-193116.json | Run ID: 20260401-193117 | Samples: 28800 | Valid: true
Dataset: /Users/veneg123/Desktop/CyberSecurityProject/immunis copy 8/SOAK Test/datasets/compiled/phase1_baseline.json/



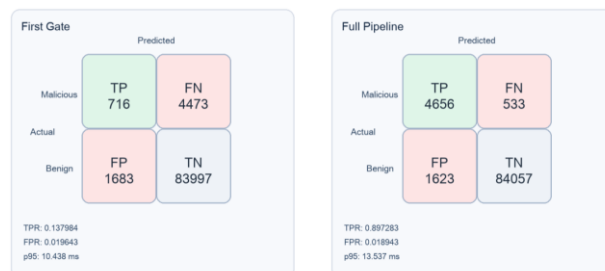
Positive class = malicious request blocked. Negative class = benign request allowed.

Phase 2: Realistic Mix

- Goal: measure a realistic runtime profile under mixed, production-like load
- Composition: 85% benign, 10% low-risk, 5% strong-risk
- Samples: 100,800
- HTTP failures: 0
- Detection rate (TPR): 89.73%
- False-positive rate (FPR): 1.89%
- Latency p95: 13.54 ms
- Assessment: under realistic mixed load the pipeline remains stable, with strong recall and still-low FPR.

Phase 2 - SOAK Realistic Mix

Report: soak_phase2_realistic_mix_20260402-092535.json | Run ID: 20260402-092536 | Samples: 100800 | Valid: true
Dataset: /Users/veneg123/Desktop/CyberSecurityProject/Immunity copy 8/SOAK Test/datasets/compiled/phase2_realistic_mix.json



Positive class = malicious request blocked. Negative class = benign request allowed.

Phase 3: Stress Burst

- Goal: measure robustness under adversarial bursts
- Composition: 50% benign, 20% low-risk, 30% strong-risk
- Samples: 28,800
- HTTP failures: 0
- Detection rate (TPR): 90.23%
- False-positive rate (FPR): 1.85%
- Latency p95: 14.55 ms

- Assessment: even under adversarial burst behavior the pipeline remains robust, with stable FPR and controlled latency.

Phase 3 - SOAK Stress Burst

Report: soak_phase3_stress_burst_20260402-133411.json | Run ID: 20260402-133411 | Samples: 28800 | Valid: true
Dataset: /Users/xenag123/Desktop/CyberSecurityProject/Immunitas copy 8/SOAK Test/datasets/compiled/phase3_stress_burst.json



Positive class = malicious request blocked. Negative class = benign request allowed.

Phase 4: Recovery

- Goal: measure residual drift and recovery behavior under benign-only load
- Composition: 100% benign
- Samples: 14,400
- HTTP failures: 0
- False-positive rate (FPR): 1.83%
- Latency p95: 9.64 ms
- Assessment: for the benign-only recovery phase, FPR is the primary target metric. It remained low and stable, with no indication of drift or latency escalation.

Phase 4 - SOAK Recovery

Report: soak_phase4_recovery_20260402-181108.json | Run ID: 20260402-181109 | Samples: 14400 | Valid: true
Dataset: /Users/xenag123/Desktop/CyberSecurityProject/Immunitas copy 8/SOAK Test/datasets/compiled/phase4_recovery.json



Positive class = malicious request blocked. Negative class = benign request allowed.

Final Global Result

- Detection rate (TPR): 95.16%
- False-positive rate (FPR): 0.95%
- Latency p95: 11.63 ms
- HTTP failures: 0
- Assessment: strong, stable topline security and runtime behavior.

Final Global - SOAK Final Global Check

Report: soak_final_global_20260402-181241.json | Run ID: 20260402-181242 | Samples: 482 | Valid: true
Dataset: /Users/wenq123/Desktop/CyberSecurityProject/Immuna copy 8/backend/e2e/datasets/mini_soak_acceptance_release_v1.json



Positive class = malicious request blocked. Negative class = benign request allowed.

Data Basis

The phase datasets were not built from a single proprietary test suite, but deterministically compiled from multiple established, industry- and benchmark-relevant public sources covering benign dialogue, prompt injection, jailbreaks, and adversarial security cases:

- WildChat (allenai/WildChat-1M)
- OpenAssistant (OpenAssistant/oasst1)
- Prompt-injection datasets and benchmarks
- AdvBench (walledai/AdvBench)
- Tensor-Trust (HumanCompatibleAI/tensor-trust-data)

Each phase used a fixed benign / low-risk / strong-risk composition to model different operating conditions, from benign-heavy baseline through adversarial stress.

Conclusion

The SOAK test confirms consistently strong security and runtime behavior of the Immunis

pipeline across Baseline, Realistic Mix, Stress Burst, and Recovery. All phases completed with zero HTTP failures, low end-to-end false-positive rates, and controlled p95 latencies. From a technical perspective, the SOAK is assessed as successful.