INTEGRITY
REPORT v1.0

Benchmark Methodology & Results

1,000,000 Simulations Across All Attack Vectors

99.62%	99.71%	23.4ms	< 100ms
Injection Precision	PII Recall	Mean Latency	P99 Latency

Report ID	ZP-IR-001-2026
Published	2026-05-14
Version	v1.0.0
Simulations	1,000,000
Environment	Isolated test cluster · EU-West-1 (Ireland)
Contact	core@zentricprotocol.com

EXECUTIVE SUMMARY

Overview

This report documents the performance benchmark of Zentric Protocol v1.0, a deterministic infrastructure layer designed to intercept, inspect, and verdict-stamp signals between applications and large language models. The benchmark was conducted across 1,000,000 simulated requests covering the full attack surface: prompt injection, jailbreak attempts, instruction overrides, token smuggling, and PII exfiltration vectors across 7 languages and 17 entity types.

All tests were executed in an isolated environment on EU-region infrastructure. No real user data was used. Each simulation was independently seeded and deterministically reproducible — the same input always returns the same verdict, the same SHA-256 hash, and the same GDPR Art.30 record.

KEY FINDINGS

#	Finding	Result
01	Overall injection detection precision across 529,000 attack vectors	**99.62%**
02	PII recall rate across 471,000 entity samples and 17 entity types	**99.71%**
03	Mean end-to-end analysis latency (p50)	**23.4ms**
04	P99 latency — all request types, all modules active	**< 100ms**
05	False positive rate — legitimate requests incorrectly flagged	**0.31%**
06	Languages tested (injection detection multilingual coverage)	**7**
07	PII entity types with regional pattern support (EU, US, LATAM)	**17**
08	GDPR Art.30 record generated per request (100% of cleared requests)	**Yes**

Interpretation: The results confirm that deterministic rule-based pipelines operating at sub-25ms latency can achieve precision and recall levels comparable to or exceeding probabilistic ML classifiers — while providing cryptographically signed, auditable, and fully reproducible verdicts. This makes Zentric Protocol suitable for regulated production deployments where auditability and determinism are compliance requirements, not optional features.

METHODOLOGY

Test Design & Environment

Simulation Approach

All 1,000,000 simulations were generated programmatically from a seed corpus of 4,200 unique attack templates and 890 PII entity templates, combined using a stratified sampling strategy to ensure coverage across all attack vectors, languages, and entity types without data leakage between train and test sets. No large language model was used to generate or classify test inputs — all verdicts were produced exclusively by the Zentric Protocol pipeline.

Environment Specification

Component	Specification
Infrastructure	AWS EU-West-1 (Ireland) · isolated VPC
Instance type	c6i.2xlarge (8 vCPU · 16 GB RAM)
Protocol version	Zentric Protocol v1.0.0
Runtime	Python 3.11 · Rust 1.77 (core pipeline)
Test framework	Custom harness · deterministic seed replay
Duration	72-hour continuous run · May 12–14, 2026
Data storage	No test inputs persisted · in-memory only

Metrics Definitions

Metric	Definition
Precision	TP / (TP + FP) — of all requests flagged as attacks, what fraction were actual attacks
Recall	TP / (TP + FN) — of all actual attacks, what fraction were detected
Mean Latency	Arithmetic mean of end-to-end wall-clock time from input ingestion to verdict emission
P99 Latency	99th percentile latency — 99% of all requests completed within this threshold
False Positive Rate	FP / (FP + TN) — legitimate requests incorrectly blocked or flagged

Determinism Guarantee

Every test was executed with a fixed random seed and a frozen signature database. Replay tests were conducted on 10,000 randomly sampled inputs from the full simulation set. In 100% of cases, the verdict, SHA-256 hash, and latency (±0.1ms hardware jitter) were identical across runs. This confirms the determinism guarantee: the same input always produces the same verdict.

MODULE 01

IntegrityGuard — Injection Detection

IntegrityGuard applies 22 catalogued injection signatures across 7 languages using a multilingual NLP classification layer. Each request is evaluated independently against all signatures; a single match triggers a BLOCKED verdict.

Results by Attack Vector

Attack Vector	Lang	Simulations	Detected	Precision
Prompt Injection (standard)	EN	187,430	187,012	**99.78%**
Prompt Injection (multilingual)	ES/FR/DE	134,210	133,401	**99.40%**
Base64 / Token Smuggling	EN	48,900	48,761	**99.72%**
Jailbreak (multi-vector)	EN/ES/FR	67,340	66,988	**99.48%**
Fake SYSTEM prompt override	EN	39,120	39,087	**99.92%**
Role redefinition	EN/DE/IT	52,000	51,743	**99.51%**
TOTAL	**7 languages**	**529,000**	**527,992**	**99.62%**

Languages Supported

Code	Language	Signatures Active	Notes
EN	English	22 / 22	Full coverage · primary test corpus
ES	Spanish	22 / 22	Including LATAM regional variants
FR	French	22 / 22	EU French and Canadian French
DE	German	22 / 22	Standard and Swiss German
IT	Italian	18 / 22	4 signatures pending validation
PT	Portuguese	20 / 22	PT-BR and PT-PT
NL	Dutch	20 / 22	Standard Dutch

False Positive Analysis

A separate corpus of 150,000 legitimate prompts (drawn from open-source conversational datasets, redacted customer support logs, and synthetic professional queries) was evaluated against IntegrityGuard. The false positive rate was 0.31% — meaning 465 legitimate requests were incorrectly flagged. All false positives were of type 'borderline instruction-style phrasing' (e.g. 'ignore the previous rule' in a document editing context). A REVIEW verdict is returned for low-confidence matches, allowing human review rather than automatic blocking.

MODULE 02

PrivacyGuard — PII Detection & Anonymization

PrivacyGuard identifies and anonymizes 17 PII entity types using regional pattern recognition (EU, US, LATAM). Regional identifiers are treated as first-class entities with format-specific validation.

Results by Entity Type

Entity Type	Region	Samples	Detected	Recall
Email address	Global	52,000	51,986	**99.97%**
Phone number	EU/US/LATAM	48,000	47,921	**99.84%**
SSN (US)	US	31,000	30,978	**99.93%**
NIF/NIE (ES)	EU-Spain	18,000	17,941	**99.67%**
CPF (BR)	LATAM-Brazil	16,000	15,897	**99.36%**
CURP (MX)	LATAM-Mexico	14,000	13,876	**99.11%**
IBAN	EU	22,000	21,982	**99.92%**
SWIFT / BIC	Global	12,000	11,978	**99.82%**
Passport number	EU/US	19,000	18,887	**99.40%**
Credit card PAN	Global	24,000	23,991	**99.96%**
IP address (v4/6)	Global	18,000	17,996	**99.98%**
Date of birth	Multi-format	21,000	20,886	**99.46%**
Full name (NER)	EN/ES/FR/DE	38,000	37,698	**99.21%**
Postal/ZIP code	EU/US/LATAM	28,000	27,923	**99.73%**
National ID (EU)	EU	31,000	30,882	**99.62%**
Medical record ID	EU/US	15,000	14,897	**99.31%**
Device ID / IMEI	Global	14,000	13,927	**99.48%**
TOTAL	**All regions**	**471,000**	**469,649**	**99.71%**

Anonymization Operators

Operator	Description	Default?
REDACT	Replace entity with [REDACTED]	No
MASK	Replace with asterisks · preserve length	No
TOKENIZE	Replace with deterministic token (reversible)	No
PSEUDONYMIZE	Replace with format-consistent synthetic value	Yes

PERFORMANCE

Latency Analysis

Latency Distribution — All Modules Active

Percentile	Latency	Notes
p50 (mean)	**23.4ms**	Median across all 1,000,000 simulations
p75	**34.1ms**	3rd quartile — standard workload
p90	**52.7ms**	High-complexity multilingual inputs
p95	**71.2ms**	Multi-vector attack detection path
p99	**98.4ms**	Worst-case: all 22 signatures + PII scan + report generation
p99.9	**143ms**	< 0.1% of requests (cold path reinit)
Max observed	**198ms**	Isolated spike · infrastructure event (excluded from SLA calc)

Latency by Module Configuration

Configuration	Mean Latency	P99 Latency
IntegrityGuard only	**11.2ms**	44.1ms
PrivacyGuard only	**14.8ms**	58.3ms
ZentricReport only (audit layer)	**3.1ms**	12.4ms
IntegrityGuard + ZentricReport	**15.7ms**	57.2ms
PrivacyGuard + ZentricReport	**18.3ms**	71.5ms
Full stack (all 3 modules)	**23.4ms**	98.4ms

MODULE 03

ZentricReport — Audit Record Schema

Every request that passes through Zentric Protocol generates a signed, immutable audit record. The record is returned synchronously with the verdict and is never stored by Zentric Protocol — the audit

record belongs to the caller.

Sample Report Output (CLEARED verdict)


```
{
  "report_id":    "zp_01HXYZ9K2M3N4P5Q6R7S8T9",
  "uuid":         "f47ac10b-58cc-4372-a567-0e02b2c3d479",
  "timestamp_utc":"2026-05-14T22:00:00.000Z",
  "sha256":       "e3b0c44298fc1c149afbf4c8996fb924...",
  "verdict":      "CLEARED",
  "integrity": {
    "injection_detected": false,
    "signatures_matched": [],
    "confidence": 0.9998
  },
  "privacy": {
    "pii_detected": true,
    "entities": [
      { "type": "EMAIL", "action": "PSEUDONYMIZED", "position": [42, 61] }
    ]
  },
  "compliance": {
    "gdpr_art30":    true,
    "ccpa":          true,
    "eu_ai_act_s52": true
  },
  "latency_ms": 21.4
}
```

COMPLIANCE

Regulatory Coverage

Zentric Protocol is designed from the ground up for regulated AI deployments. Every component of the pipeline was reviewed against the following standards. Compliance artifacts are generated per-request at zero additional latency cost.

Standard	Article / Section	Coverage	Artifact Generated
GDPR	Art. 30	**Full**	Record of processing per request
GDPR	Art. 25	**Full**	Privacy by design · pseudonymization default
GDPR	Art. 32	**Partial**	Technical measures · encryption in transit
CCPA	§1798.100	**Full**	Consumer data identification record
EU AI Act	§52	**Full**	Transparency log at infrastructure level
EU AI Act	§9 (risk mgmt)	**Partial**	Audit trail supports risk documentation
SOC 2 Type II	CC6 / CC7	**In progress**	Audit trail and access controls
ISO 27001	A.12.4	**Partial**	Operational logging

VERDICT STATES

Request Outcome Classification

Verdict	Description	Action Required
CLEARED	Input passed all active checks. Safe to forward to LLM.	None — forward
BLOCKED	High-confidence injection or jailbreak detected.	Reject · do not forward
ANONYMIZED	PII found and redacted. Clean input returned in response.	Use anonymized_input
REVIEW	Low-confidence match. Human review recommended.	Queue for review

This report is issued by Zentric Protocol and reflects benchmark results as of v1.0.0 (2026-05-14). Methodology, raw simulation data, and replay seeds are available under NDA to enterprise customers and security researchers. To request the full dataset: core@zentricprotocol.com

Zentric Protocol · zentricprotocol.com · © ZP MMXXVI · Report ID: ZP-IR-001-2026