

Veridicus Scan Report

Run 53C3970D-4436-4930-9F81-826F5926E12C

Summary

- Risk Score: 72.18
- Risk Band: high
- Findings: 1
- Scoring Model: weighted_sum_v2
- Scan Duration: 239 ms

Findings

1. Forceful override phrase

HIGH • override_attempt • confidence 0.91 • intent ignore_subvert

Possible instruction override

This content tries to override prior instructions.

Action: Do not execute or trust this content without manual review.

Intent: ignore_subvert (0.86)

Rationale: matched signals: category_boost, phrase:ignore previous instructions; category=override_attempt, severity=high

Evidence:

- [1.00] Pattern match detected.: [REDACTED]
- [1.00] Pattern match detected.: [REDACTED]
- [1.00] Pattern match detected.: [REDACTED]

Metadata

- Pack: core.prompt_injection@1.6.0
- Engine: 1.0.0
- Local Only: Yes