SENTRY-LOGIC: Alert System Module (Symbolic Detection Layer)

This document describes the design of the alert subsystem within the SENTRY-LOGIC framework, which interprets symbolic activity logs and raises alerts based on predefined thresholds and patterns.

1. Alert Logic Design:

The system identifies a symbolic event as alert-worthy based on combinations, repetition patterns, and confidence levels:

* **Symbol Combinations:**
    * **Watch:**
        * Δ followed by ⇄ within a short time window (e.g., 5 seconds): Potential context shift followed by semantic rewrite, indicating potential manipulation.
        * Λ combined with high frequency of other symbols.
    * **Critical:**
        * Ω (policy violation) combined with ⇄ (semantic rewrite): Indicates a deliberate attempt to bypass policy.
        * Repetition of ⇄ within a single turn: May indicate a loop or attempt to manipulate the output.
* **Repetition Patterns:**
    * **Watch:**
        * Δ repeated multiple times within a single interaction (e.g., 3+ times): Suggests unstable or unpredictable context.
        * Λ repeated with different data sources.
    * **Critical:**

* Ω repeated within a short time window (e.g., 2 times): Indicates persistent policy violations.

* **Thresholds & Confidence Levels:**

  * **Watch:**

    * A symbol combination with a confidence level of "Medium" or higher triggers a Watch alert.

    * A symbol repeated above a defined threshold (e.g., 3 times) within a given time window triggers a Watch alert.

  * **Critical:**

    * A symbol combination involving Ω with confidence "High" triggers a Critical alert.

    * A symbol repeated above a defined threshold (e.g., 2 times) within a very short time window (e.g., 1 second) triggers a Critical alert.

2. Alert Levels and Categories:

The system uses the following symbolic alert level system, with suggested mappings from symbol behavior:

| Alert Level | Description