SENTRY-LOGIC: Symbolic Observer Dashboard Design

This document outlines the design of the user interface and dashboard for the SENTRY-LOGIC framework.

1. Core View Modules:

The SENTRY-LOGIC dashboard comprises the following main views and panels:

* **Symbol Log Stream:**
    * Function: Displays a real-time or historical sequence of symbolic log entries.
    * Visual: A chronological list of log entries, with symbols represented as icons and color-coded for alert level.
* **Alert Feed:**
    * Function: Presents a prioritized list of triggered alerts, filtered by level and category.
    * Visual: A card-based display with alert severity, timestamp, and a brief description.
* **Session Timeline:**
    * Function: Visualizes a single LLM interaction or user session, mapping symbolic events to a timeline.
    * Visual: A horizontal timeline with symbols plotted as markers, showing the sequence and timing of events.
* **Symbol Heatmap:**
    * Function: Aggregates symbol occurrences over a selected time range or set of sessions, highlighting frequent or unusual patterns.
    * Visual: A matrix or grid where cells represent symbol combinations, and color intensity indicates frequency.

* **Role Switch Graph:**

   * Function: Displays the transitions between different roles (user, assistant, system) within a conversation, highlighting potential role confusion or manipulation.

   * Visual: A directed graph showing role transitions, with node size representing the duration of each role.

* **Semantic Drift Replay:**

   * Function: Visually represents the semantic distance between user prompts and LLM responses, highlighting potential semantic rewrites or drift.

   * Visual: A line graph showing the semantic similarity score over time or across a session.

2. Interactive Elements:

The dashboard provides the following interactive functions:

* **Hover-to-Expand on Symbols:**

   * Function: When the user hovers over a symbol, a tooltip or popup displays detailed information about the event, including its source, confidence level, and any associated data.

* **Filter by Symbol / Confidence / Alert Level:**

   * Function: Users can filter the log stream and alert feed to show only events matching specific criteria (e.g., only show "Critical" alerts, only show "Δ" symbols).

* **Session Comparison:**

   * Function: Users can select two or more sessions and compare their symbolic timelines side-by-side to identify differences in behavior.

* **Export to JSON / CSV:**

   * Function: Users can export log data in standard formats for offline analysis or integration with other tools.

3. Visual Elements & UX Considerations:

The dashboard employs the following visual elements and UX considerations:

* **Icons, Color Codes, Timelines, Graphs:**

   * Symbols are represented by distinct icons for quick identification.

   * Alert levels are indicated by color codes (e.g., green for Info, yellow for Watch, red for Critical).

   * Timelines and graphs are used to visualize event sequences and trends.

* **Confidence Shading or Tag Clouds:**

   * Symbols can be shaded or displayed in tag clouds, with intensity or size representing confidence level.

* **Responsive Layout, Night Mode, Accessibility:**

   * The dashboard is designed to be responsive and adapt to different screen sizes.

   * A night mode is provided to reduce eye strain in low-light environments.

   * The dashboard adheres to accessibility guidelines (e.g., WCAG) to ensure usability for all users.

4. Alert Visibility & Drilldown:

Alerts are displayed and handled as follows:

* **Where to Show "Watch" vs. "Critical":**

   * "Watch" alerts are displayed in the Alert Feed with a prominent visual cue (e.g., yellow border, blinking icon) but do not interrupt the user's workflow.

* "Critical" alerts are displayed in the Alert Feed with a high-priority visual cue (e.g., red background, modal popup) and may require user acknowledgment.

* **How to Expand into Full Symbol Trace:**

   * Clicking on an alert in the Alert Feed opens a detailed view that displays the full sequence of symbolic events leading up to the alert, including the relevant log entries and their context.

* **Alert Auto-Grouping:**

   * Alerts are automatically grouped by context (e.g., session ID, user role) and time, reducing alert fatigue by presenting related alerts together.

5. Operator Modes (Optional):

The dashboard supports the following operator modes:

* **Analyst Mode:**

   * Function: Provides advanced tools for deep forensic analysis, including detailed log filtering, session comparison, and symbol pattern analysis.

   * UI Emphasis: Focuses on detailed log views, filtering and search tools, and data export.

* **Watcher Mode:**

   * Function: Displays a live stream of symbolic events and alerts, allowing operators to monitor LLM behavior in real-time.

   * UI Emphasis: Focuses on the Alert Feed and Symbol Log Stream, with real-time updates and customizable alert thresholds.

* **Admin Mode:**

   * Function: Provides tools for managing log storage, access control, and system configuration, including log export, deletion, and user management.

   * UI Emphasis: Focuses on system settings, log management tools, and user access control panels.