

## Contents

[9.8] Human-AI Team Dysfunction . . . . .	1
[ . . . . .]	2

### [9.8] Human-AI Team Dysfunction

**1. Operational Definition:** The breakdown of effective collaboration, communication, and role clarity between human security team members and AI agents, leading to misaligned actions, duplicated effort, or critical tasks being dropped.

**2. Main Metric & Algorithm:**

- **Metric:** Task Conflict Rate (TCR). Formula:  $TCR = N_{conflicting\_actions} / N_{total\_AI\_human\_action\_pairs}$ .

- **Pseudocode:**

```
python

def calculate_tcr(ai_actions, human_actions, start_date, end_date):
    # Get actions on the same alert within a short time window
    conflicting_pairs = []
    time_window = timedelta(minutes=5)

    for h_action in human_actions:
        for a_action in ai_actions:
            if (h_action.alert_id == a_action.alert_id and
                abs(h_action.timestamp - a_action.timestamp) < time_window and
                h_action.action != a_action.action):
                conflicting_pairs.append((h_action, a_action))

    # Estimate total potential pairs for a ratio
    total_potential_pairs = ... # Complex calculation based on overlapping assignments and
    # A simpler proxy: Total number of alerts worked by both AI and humans

    N_conflicts = len(conflicting_pairs)
    # Using a simpler proxy denominator
    N_alerts_worked = count_alerts_worked_by_both(start_date, end_date)

    if N_alerts_worked > 0:
        TCR = N_conflicts / N_alerts_worked
    else:
        TCR = 0

    return TCR
```

- **Alert Threshold:**  $TCR > 0.1$  (Conflicting actions occur on more than 10% of alerts worked by both).

**3. Digital Data Sources (Algorithm Input):**

- **SOAR/SIEM APIs:** Detailed logs of all actions taken by AI agents and human analysts (`actor_type`, `alert_id`, `action`, `timestamp`).
- 4. Human-To-Human Audit Protocol:** Run a workshop simulating an incident response. Include the AI as a team member. Observe the workflow: Do humans know what the AI is doing? Does the AI undermine human commands? Is there confusion about roles?

## **5. Recommended Mitigation Actions:**

- **Technical/Digital Mitigation:** Improve the AI’s “teamplay” by having it broadcast its intentions and actions clearly via a dedicated notification channel (e.g., “I am prioritizing alert X”).
  - **Human/Organizational Mitigation:** Clearly define the role of the AI in team charters (e.g., “The AI is an assistant for triage, not for final decision-making”).
  - **Process Mitigation:** Design clear playbooks that specify which agent (human or AI) is responsible for which specific action in a given scenario to prevent overlap and conflict.
- 

[