

Contents

[5.10] Mental Model Confusion	1
-----------------------------------------	---

[5.10] Mental Model Confusion

1. Operational Definition: A situation where an analyst holds an incorrect or incomplete internal understanding of how a system, tool, or attack works, leading them to misinterpret data and draw erroneous conclusions.

2. Main Metric & Algorithm:

- **Metric:** Hypothesis Invalidation Rate (HIR). Formula: $HIR = (\text{Number of times an analyst's initial hypothesis about an alert is proven wrong by subsequent evidence}) / (\text{Total number of alerts where they formed a hypothesis})$.

- **Pseudocode:**

```
python

def calculate_hir(tickets, analyst_id):
    # Get tickets where the analyst was the primary investigator
    analyst_tickets = get_tickets(primary_analyst=analyst_id, has_initial_assessment=True)
    invalidation_count = 0

    for ticket in analyst_tickets:
        initial_hypothesis = ticket.initial_assessment # e.g., "This is a FP"
        final_verdict = ticket.final_verdict           # e.g., "True Positive - Malware"
        if initial_hypothesis != final_verdict:
            invalidation_count += 1

    return invalidation_count / len(analyst_tickets)
```

- **Alert Threshold:** $HIR > 0.4$ (The analyst's initial hypothesis is wrong more than 40% of the time, indicating a flawed mental model).

3. Digital Data Sources (Algorithm Input):

- **Ticketing System (Jira/ServiceNow):** Requires custom fields **Initial Hypothesis** and **Final Verdict** to be populated by analysts and validated by a senior analyst or automated tool.

4. Human-to-Human Audit Protocol: During an incident retrospective, ask the analyst: “Walk me through your thought process when you first saw this alert. What did you think was happening and why?” Then compare this to the evidence. Probe for misunderstandings about tool functionality (e.g., “I thought the firewall would have blocked this”) or attack techniques.

5. Recommended Mitigation Actions:

- **Technical/Digital Mitigation:** Implement a “mental model” wiki that documents common systems and attacks, explicitly addressing frequent misconceptions.
- **Human/Organizational Mitigation:** Institute a robust mentorship program where junior analysts regularly review cases with senior analysts to correct and refine their mental models.

- **Process Mitigation:** Create a culture of “blameless post-mortems” that focus on identifying and correcting flawed mental models rather than assigning fault for errors.