

## Contents

[9.3] Paradosso dell'Avversione agli Algoritmi . . . . . 1

### [9.3] Paradosso dell'Avversione agli Algoritmi

**1. Definizione Operativa:** La tendenza controllistica di diffidare o ignorare sistemi IA/automatizzati accurati e affidabili, spesso dopo aver sperimentato un piccolo numero di errori precedenti, portando a un sottoutilizzo di uno strumento prezioso.

#### 2. Metrica Principale e Algoritmo:

- **Metrica:** Indice di Mancato Utilizzo del Sistema (SDI). Formula:  $SDI = (N_{azioni\_intraprese\_senza\_co}) / N_{azioni\_totali}$  per compiti dove il supporto IA è disponibile e consigliato.
- **Pseudocodice:**

```
def calculate_sdi(analyst_actions, ai_capable_alerts, start_date, end_date):  
    # Ottenere tutte le azioni intraprese dagli analisti su avvisi che l'IA avrebbe potuto eseguire  
    relevant_actions = [  
        a for a in analyst_actions  
        if a.alert_id in ai_capable_alerts  
        and a.timestamp between start_date and end_date  
    ]  
  
    # Verificare se l'analista ha consultato l'IA prima di intraprendere l'azione  
    # Questo richiede log dal sistema IA che mostra eventi di query/accesso  
    actions_without_ai = [  
        a for a in relevant_actions  
        if not exists_ai_access_log(a.alert_id, a.analyst_id, a.timestamp)  
    ]  
  
    N_total = len(relevant_actions)  
    N_disused = len(actions_without_ai)  
  
    if N_total > 0:  
        SDI = N_disused / N_total  
    else:  
        SDI = 0  
  
    return SDI
```

- **Soglia di Avviso:**  $SDI > 0.75$  (Gli analisti non stanno utilizzando il supporto IA disponibile per oltre il 75% delle azioni rilevanti).

#### 3. Fonti Dati Digitali (Input dell'Algoritmo):

- **API di SOAR/SIEM:** Log di tutte le azioni intraprese dagli analisti su avvisi (`analyst_id`, `alert_id`, `action`, `timestamp`).
- **Log di Audit del Sistema IA:** Record di ogni volta che un analista interroga l'IA per una raccomandazione su uno specifico avviso (`analyst_id`, `alert_id`, `query_timestamp`).

- **CMDB/Database Risorse:** Un elenco di `alert_types` o `asset_classes` per cui il supporto IA è disponibile e consigliato (`ai_capable_alerts`).
- 4. Protocollo di Audit Umano-Umano:** Intervistare i membri del team: “Puoi descrivere un momento in cui sei stato in disaccordo con il suggerimento dello strumento IA? Che cosa hai fatto? Quanto spesso controlli le sue raccomandazioni?” Cercare narrazioni di sfiducia basate su eventi passati isolati piuttosto che su metriche di performance attuali.

**5. Azioni di Mitigazione Consigliate:**

- **Mitigazione Tecnica/Digitale:** Migliorare la trasparenza del sistema IA fornendo spiegazioni concise e comprensibili per le sue raccomandazioni (XAI - IA Spiegabile).
- **Mitigazione Umana/Organizzativa:** Mostrare l'accuratezza complessiva dell'IA e il suo valore nelle riunioni di team, affrontando il “perché” dietro alle sue raccomandazioni per ricostruire la fiducia.
- **Mitigazione di Processo:** Inizialmente rendere la consultazione dell'IA un passo obbligatorio nel playbook per tipi di avviso specifici, forzando il rimpegno per dimostrare affidabilità migliorata.