

Contents

[9.4] Trasferimento di Autorità dell'IA 1

[9.4] Trasferimento di Autorità dell'IA

1. Definizione Operativa: La cessione inconscia dell'autorità di processo decisionale e della responsabilità al sistema IA, diminuendo il senso di agency dell'analista e l'impegno critico con il processo di sicurezza.

2. Metrica Principale e Algoritmo:

- **Metrica:** Punteggio di Perdita di Agency (ALS). Formula: $ALS = \frac{N_{raccomandazioni_IA_accettate_lette}}{N_{raccomandazioni_IA_totali_accettate}}$.
 - **Pseudocodice:**

```
def calculate_als(ai_recommendations, analyst_actions, start_date, end_date):
    # Ottenere tutte le raccomandazioni IA che sono state accettate da un analista
    accepted_recommendations = [
        r for r in ai_recommendations
        if exists_analyst_action(r.alert_id, r.recommended_action)
        and r.timestamp between start_date and end_date
    ]

    # Filtrare per quelle accettate senza alcuna modifica o nota aggiuntiva
    verbatim_acceptances = [
        r for r in accepted_recommendations
        if not exists_analyst_notes(r.alert_id)    # Nessun ragionamento aggiuntivo fornito
        and not was_modified(r.alert_id, r.recommended_action) # L'azione intrapresa corrisponde alla raccomandazione
    ]

    N_total_acceptances = len(accepted_recommendations)
    N_verbatim = len(verbatim_acceptances)

    if N_total_acceptances > 0:
        ALS = N_verbatim / N_total_acceptances
    else:
        ALS = 0

    return ALS
```

- **Soglia di Avviso:** ALS > 0.9 (Oltre il 90% delle raccomandazioni IA accettate vengono seguite esattamente senza modifica umana o ragionamento documentato).

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

- **API di Sistema IA e Sistema di Ticketing:** Come in 9.2 e 9.3, richiedendo log di raccomandazioni IA (`recommended_action`), azioni dell'analista (`action_taken`), e note dell'analista (`resolution notes, comments`).

4. Protocollo di Audit Umano-Umano: Osservare un analista che utilizza lo strumento IA.

Notare se scorrono oltre la raccomandazione IA per guardare i dati grezzi loro stessi. Nelle interviste, chiedere: “Chi è ultimamente responsabile se un’azione basata su una raccomandazione IA porta a una violazione?” Ascoltare le risposte che deresponsabilizzano lo strumento.

5. Azioni di Mitigazione Consigliate:

- **Mitigazione Tecnica/Digitale:** Progettare l’interfaccia utente per richiedere un campo obbligatorio “Razionale dell’Analista” da compilare prima di procedere con un’azione consigliata dall’IA su avvisi ad alta severità.
- **Mitigazione Umana/Organizzativa:** La leadership deve comunicare esplicitamente che l’analista, non lo strumento, è responsabile delle decisioni. Rinforzare questo nella formazione.
- **Mitigazione di Processo:** Implementare una politica dove una certa percentuale di decisioni guidate dall’IA viene selezionata casualmente per revisione obbligatoria tra pari.