HaiR: A Human-Centric AI Remediation Framework

Enabling Safe, Explainable, and Accessible Cybersecurity Remediation for Resource-Constrained Organisations

Dr Olayinka Adeboye
a.a.adeboye@salford.ac.uk
University of Salford, Manchester

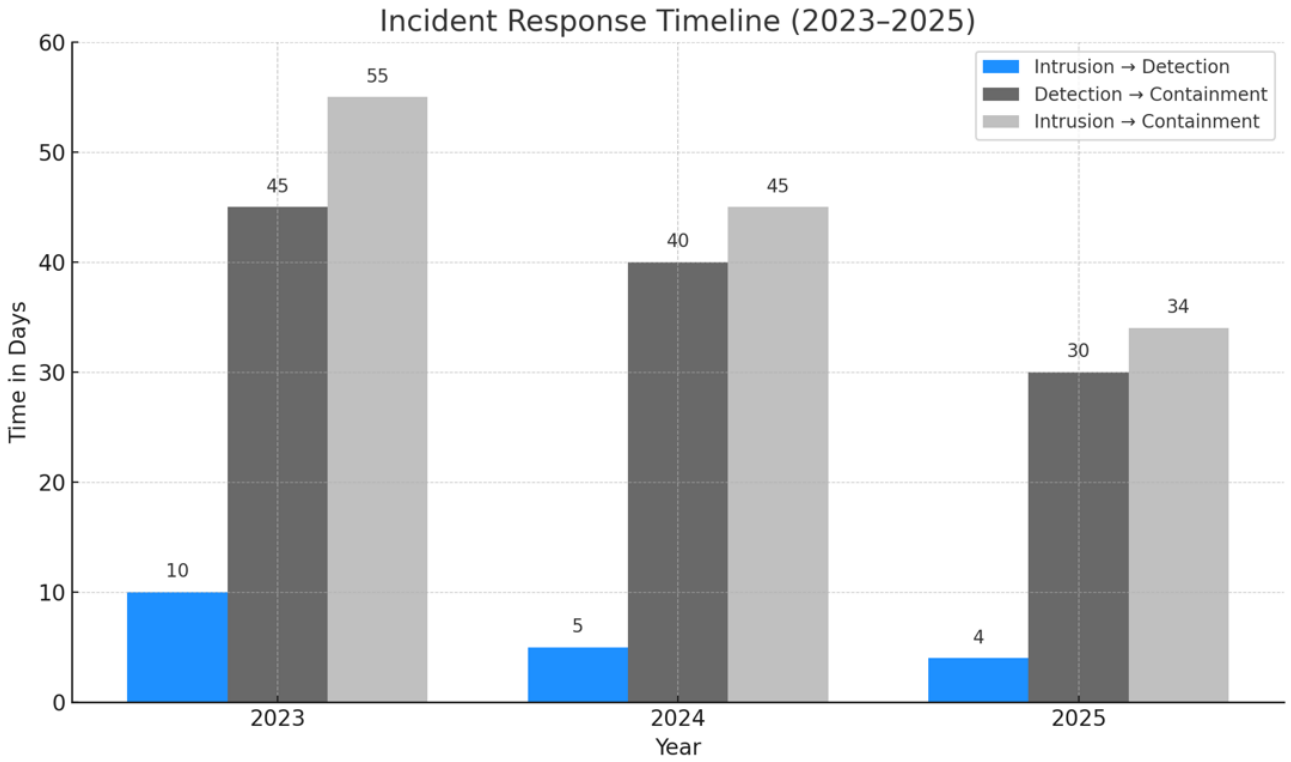
Authors:
Dr Rabab Al Zaidi
Dr Lee Speakman
Muhammad Ateeq
Aderemi Adenihun
Segun Akinseye

Motivation

Detection is not Protection

The "Time-to-Remediate" remains the single biggest gap in modern cybersecurity defense.

- **32-Day Median:** Time taken to remediate systems (Verizon 2025).
- **SME Fatigue:** Lack of experts leads to delayed response.
- **SME Challenges:** Technical & Organisational Barriers



	Incidents				Breaches			
Industry	Total	Small (1–1,000)	Large (1,000+)	Unknown	Total	Small (1–1,000)	Large (1,000+)	Unknown
Total	22,052	3,049	982	18,021	12,195	2,842	751	8,602

CURRENT RESEARCH GAPS

SOAR Limits

High complexity and cost make it unsuitable for non-expert users in SME environments.

Autonomous Risks

Fully automated AI can lead to compliance violations and critical system downtime.

LLM Constraints

Issues with logic control, reduced auditability, and incomplete system context.

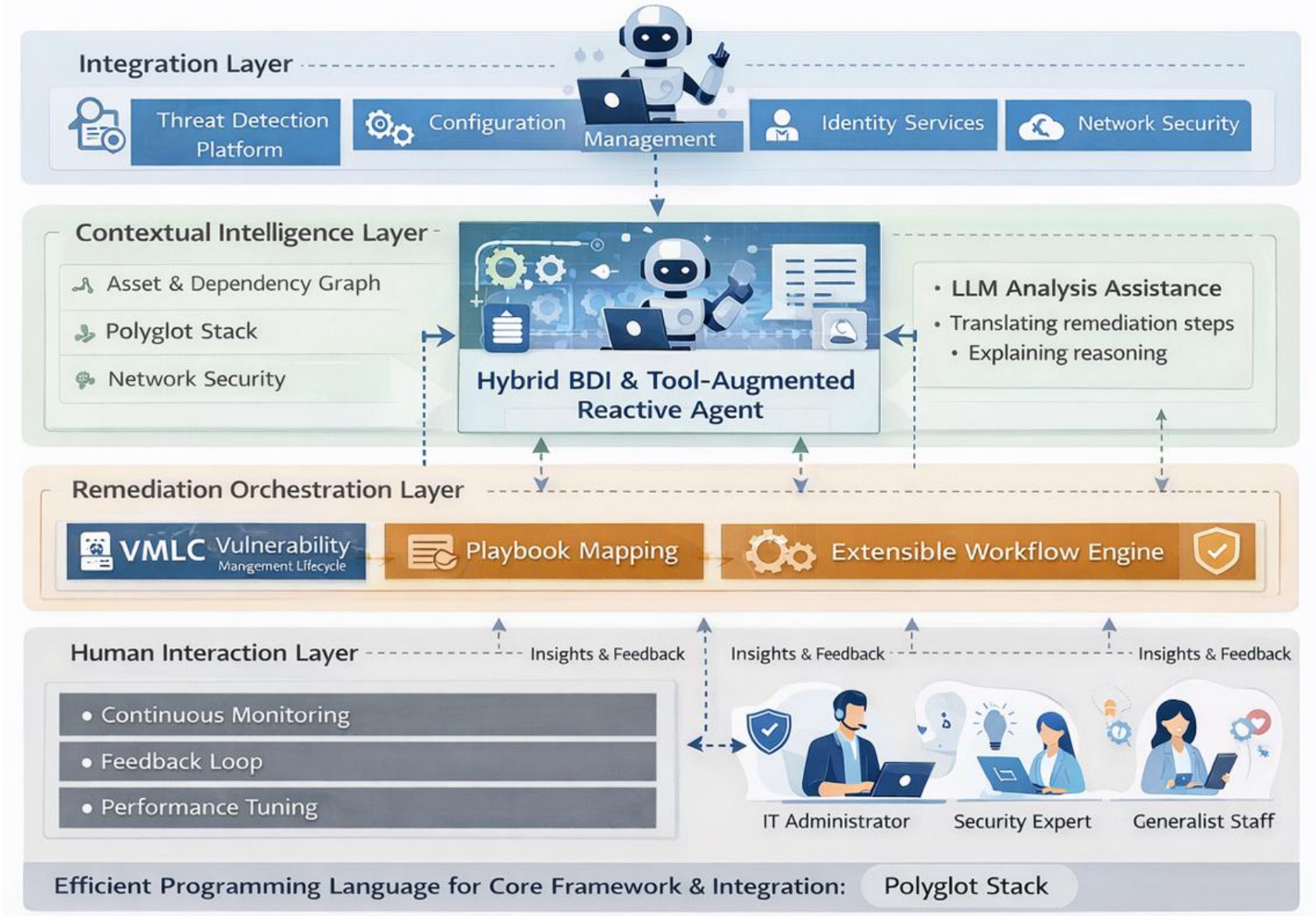
AIM & OBJECTIVES

HaiR

FRAMEWORK GOAL

- Design a **four-layer** remediation architecture.
- Develop a **hybrid agent** for contextual reasoning.
- Align actions with **incident response playbooks**.
- Ensure **auditable, policy-compliant** execution.

THE HAIR FRAMEWORK



Experimental Setup

Parameter	Experimental Configuration
Environment	Simulated Small-to-Medium Enterprise (SME) Infrastructure
Tooling	Wazuh (SIEM/XDR), OpenVAS (Vulnerability Scanner)
Scenarios	1. Misconfiguration 2. Vulnerable Package Management
Participants	2 Security Practitioners for Human-in-the-Loop Evaluation

PERFORMANCE METRICS

EVALUATION RESULTS FOR THE SCENARIOS

Scenario	Manual	SOAR-only	Proposed Framework
Misconfiguration	95 mins	35 mins	12 mins
Vulnerable package	120 mins	50 mins	18 mins

MORE EVALUATION RESULTS FOR THE SCENARIOS

Control Aspect	Manual	SOAR-only	Proposed Framework
Policy enforcement	Inconsistent	Partial	Full
Audit completeness	Partial	Partial	Complete
Rollback support	Manual	Limited	Automatic
Unsafe actions	Possible	Possible	Non observed

COMPARISONS OF LLM-ONLY AGENTS AND PROPOSED HYBRID BDI + TOOLS + POLICY

Criterion	LLM-only Agent	Proposed Hybrid (BDI + tools +policy)
Explainability	Minimal	High
Safety	Higher risk of unsafe actions	Strong (approval, audit, rollback)
Determinism	Non-deterministic outputs	Deterministic planning +bounded variability
Cost	High if API-enabled	Low
SME deployment	Harder (dependency on model hosting)	Lightweight solution for deployment (Go binary, light python; offline-capable)

RESPONSIBLE AI PRINCIPLES

- **Human Oversight:** Mandatory approval for all changes.
- **Explainability:** Natural language "Why" and "How".
- **No Full Autonomy:** Guided, never independent.
- **Reversibility:** Ability to restore system integrity.

University of Salford MANCHESTER

Thank You

Questions & Discussion

Contact: a.a.adeboye@salford.ac.uk
University of Salford