

CloudStrike AI

by: Ritvik Indupuri

1. Introduction & Core Philosophy

CloudStrike AI is a full-stack, AI-powered cybersecurity threat modeling sandbox designed for the modern cloud-native landscape. It empowers security professionals, from analysts to penetration testers, to move from a reactive to a **proactive security posture**.

The core philosophy is to create an intelligent "**sparring partner**" for defenders. Instead of relying on static threat intelligence reports or mock data, CloudStrike AI provides a dynamic, interactive environment where users can:

1. **Generate** a realistic but simulated attack based on a natural language description.
2. **Model** the full impact of that attack across a simulated cloud environment.
3. **Analyze** an AI-generated defense and iteratively **improve** it through simulated engagements.

This transforms security training and defense validation from a passive exercise into an active, hands-on feedback loop.

2. System Architecture

CloudStrike AI is built using a modern, type-safe, and AI-centric technology stack. The architecture is split between a responsive frontend and a sophisticated multi-agent AI backend. The diagram below provides a visual overview of how these components interact to deliver the platform's functionality.

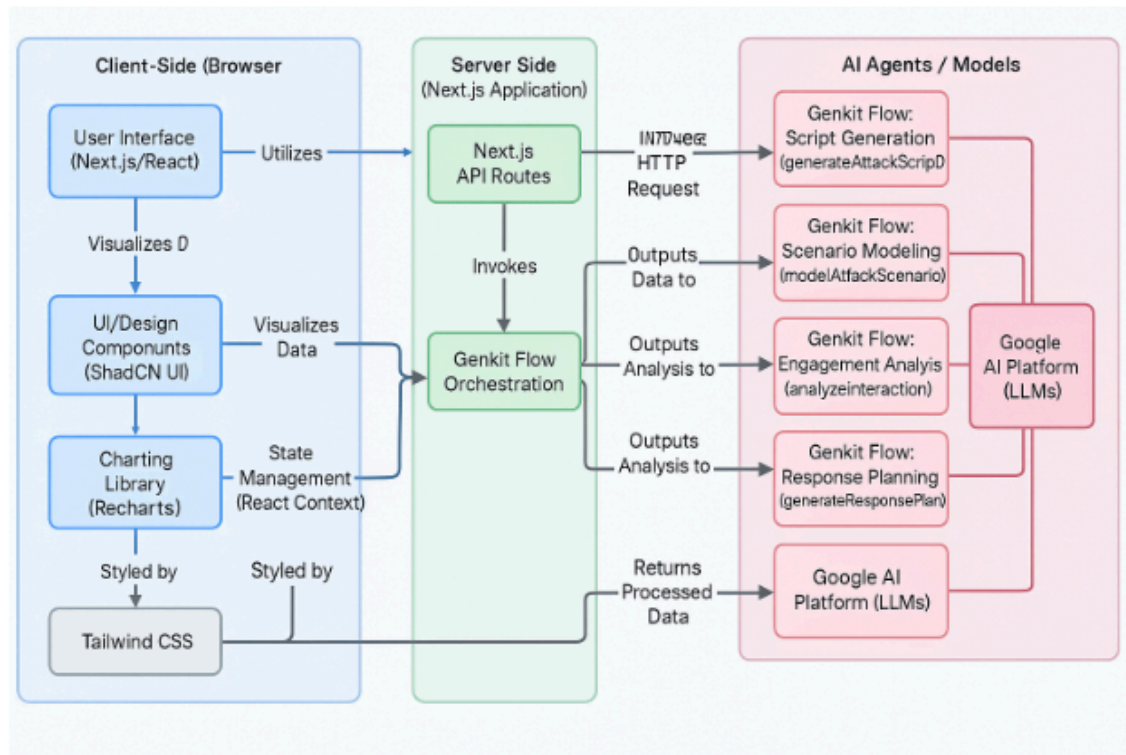


Figure 1: CloudStrike AI System Architecture

This diagram illustrates the interconnected components of the CloudStrike AI system. It details the flow from the client-side browser, which includes the user interface and charting libraries, to the server-side Next.js application. The server-side invokes a series of AI agents, orchestrated by Genkit, to handle tasks such as script generation, scenario modeling, engagement analysis, and response planning, all of which leverage Google AI Platform's large language models (LLMs).

2.1. Frontend

The frontend is built for a responsive and professional user experience, leveraging a modern technology stack.

- **Framework:** Next.js (App Router) with React 18
- **Language:** TypeScript
- **UI Components:** ShadCN UI - a collection of accessible, composable components.
- **Styling:** Tailwind CSS with a custom dark theme for high contrast and readability.
- **Data Visualization:** Recharts for creating interactive charts.

- **State Management:**
 - **React Context API:** Manages the global state of the simulation session.
 - **Persistent Session History:** Uses `localStorage` to save and reload past scenarios.

2.2. AI Backend & Core Logic

The backend is an AI-centric system orchestrated to perform complex, multi-step tasks.

- **AI Orchestration Framework:** **Google Genkit** is the core, defining and managing a series of interconnected AI "flows."
 - **AI Models:** A "model ladder" is used for resilience:
 - **Primary:** `googleai/gemini-1.5-flash` for speed and reasoning.
 - **Fallback:** `googleai/gemini-pro` for maximum uptime.
 - **Structured Data I/O:** **Zod** schemas are used to ensure the AI returns predictable, type-safe JSON.
 - **Safety & Constraints:** Genkit's `safetySettings` are configured to allow for the discussion of dangerous content in a simulated and harmless manner.
-

3. The Multi-Agent AI System: How It Works

CloudStrike AI's intelligence is orchestrated through a series of specialized AI agents (Genkit flows). Each piece of the scenario is dynamically generated, ensuring a unique simulation every time.

Flow 1: `generateAttackScript` (The "Red Team")

This flow acts as a red team expert, generating a realistic, but **simulated**, attack script.

- **Trigger:** A natural language description of an attack from the user.
- **Key Output:** A safe-to-execute script.

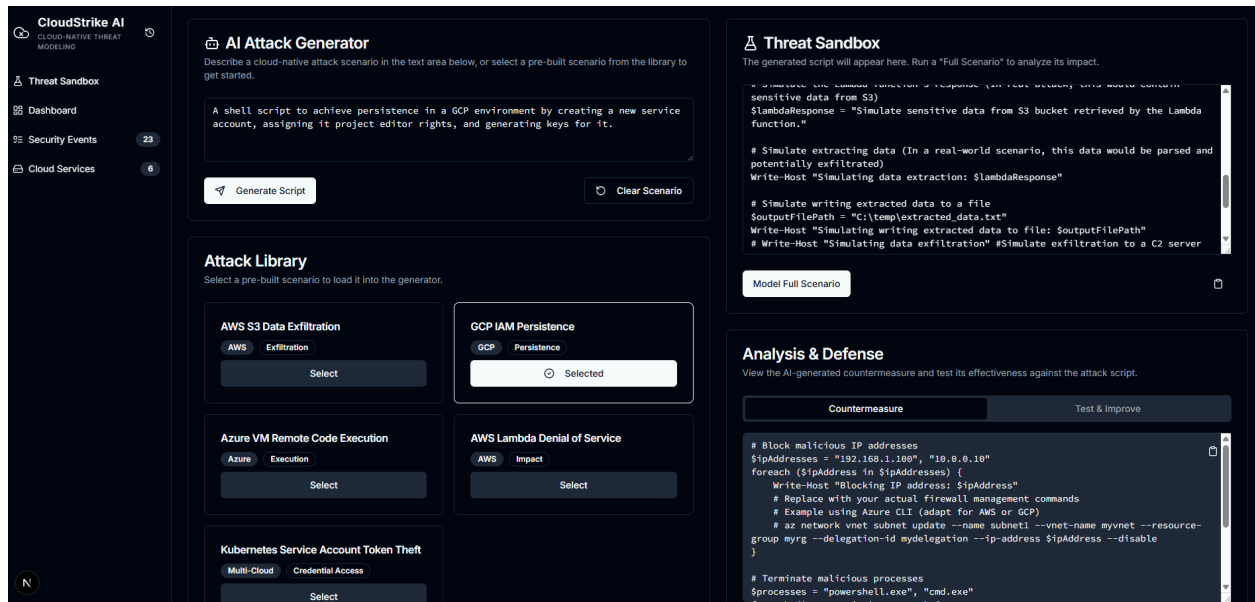


Figure 2: AI Attack Generator

This interface panel, titled "AI Attack Generator," allows users to input a natural language description or select a pre-built scenario from the "Attack Library" to generate a simulated attack script and its corresponding countermeasures.


Flow 2: **modelAttackScenario** (The "Cloud Intrusion Detection System")


This is the central orchestrator, analyzing the attack script to generate a complete, interconnected dataset representing the full impact.

- **Trigger:** The user submits a script for analysis.
- **Key Outputs:** A threat analysis report, a list of affected cloud resources, security events, dashboard metrics, and a suggested countermeasure.

Affected Cloud Resources

A list of cloud assets impacted by the simulated attack scenario.


 **webserver-01**


 EC2 Instance in us-east-1

Resource ID:
i-8abcdef1234567890

Status:
Compromised

Reason:
Status is 'Compromised' because the script successfully executed commands on this EC2 instance.


 **database-instance**


 RDS in us-east-1

Resource ID:
arn:aws:rds:us-east-1:123456789012:db:mysqlinstance

Status:
Vulnerable

Reason:
Status is 'Vulnerable' because the script attempted to access this RDS instance using weak credentials.


 **log-bucket**


 S3 in us-east-1

Resource ID:
arn:aws:s3::my-log-bucket

Status:
Compromised

Reason:
Status is 'Compromised' because the script uploaded malicious log files to this S3 bucket.


 **lambda-function**


 Lambda in us-east-1

Resource ID:
arn:aws:lambda:us-east-1:123456789012:function:my-lambda

Status:
Compromised

Reason:
Status is 'Compromised' because the script triggered this Lambda function, which executed malicious code.

 **user-data**

 DynamoDB in us-east-1

Resource ID:
arn:aws:dynamodb:us-east-1:123456789012:table/user-data

Status:
Compromised

Reason:
Status is 'Compromised' because the script accessed and downloaded sensitive user data from this DynamoDB table.

Figure 3: Affected Cloud Resources

The "Affected Cloud Resources" section, dynamically generated based on the simulated attack script, lists various cloud assets and their security statuses ("Compromised," "Vulnerable"), along with concise explanations.

| Security Events | | | | | |
|---|------------------------|----------|--|-----------------|---------|
| A log of all security incidents detected during the simulation. | | | | | |
| Event ID | Timestamp | Severity | Description | Status | Actions |
| EVT-001 | 7/26/2024, 10:30:00 AM | Critical | Investigating: Suspicious AWS access key usage detected. [T1071.001] | Investigating | Respond |
| EVT-002 | 7/26/2024, 10:31:00 AM | Critical | Investigating: Unauthorized EC2 instance startup detected. [T1574.001] | Investigating | Respond |
| EVT-003 | 7/26/2024, 10:32:00 AM | High | Investigating: Unusual network traffic detected from compromised EC2 instance. [T1040.003] | Investigating | Respond |
| EVT-004 | 7/26/2024, 10:33:00 AM | High | Investigating: Failed login attempts to AWS RDS Instance. [T1078] | Investigating | Respond |
| EVT-005 | 7/26/2024, 10:34:00 AM | Critical | Investigating: Data exfiltration attempt detected from compromised S3 bucket. [T1567.001] | Investigating | Respond |
| EVT-006 | 7/26/2024, 10:35:00 AM | Critical | Action Required: AWS Lambda function invoked without authorization. [T1566.001] | Action Required | Respond |
| EVT-007 | 7/26/2024, 10:36:00 AM | High | Investigating: Malicious process detected on compromised EC2 instance. | Investigating | Respond |
| EVT-008 | 7/26/2024, 10:37:00 AM | Critical | Action Required: Suspicious AWS API calls detected from unknown IP address. | Action Required | Respond |
| EVT-009 | 7/26/2024, 10:38:00 AM | Critical | Investigating: Attempt to access sensitive data in DynamoDB table. | Investigating | Respond |
| EVT-010 | 7/26/2024, 10:39:00 AM | High | Investigating: Unusual file activity observed on compromised EC2 instance. [T1071.001] | Investigating | Respond |

Figure 4: Security Events Log

A comprehensive log titled "Security Events" displays all security incidents detected during the simulation. Each event includes a unique ID, timestamp, severity, a description, and its current status.

Flow 3: **analyzeInteraction** (The "Purple Team Lead")

This flow simulates a head-to-head engagement between the attack script and the defense script, embodying the "sparring partner" concept.

- **Trigger:** The user decides to test the generated countermeasure.
- **Key Outputs:** A chronological interaction log, an effectiveness score, a count of **blocked attacks**, and an AI-improved defense script.

Analysis & Defense

View the AI-generated countermeasure and test its effectiveness against the attack script.

Countermeasure

Test & Improve

```
#!/bin/bash
# Rule 1: Deny root deletion commands
Deny root deletion commands
# Rule 2: Rate limit SSH login attempts
Rate limit SSH login attempts
#Enhanced Rule 3: Add additional checks to block or flag suspicious network
activity from nmap
if [[ "$PROCESS" == "nmap" ]]; then
    # Flag or block suspicious nmap processes based on further analysis or other
    available information
fi
#Enhanced Rule 4: Improve logging for all events for easier security analysis
```



AI-Generated Defense

This is the AI's suggested defense script. Click the "Test & Improve" tab to run a simulated engagement and generate an even better version.



Re-Test & Improve

Figure 5: Analysis & Defense Interface

The "Countermeasure" sub-tab presents the AI-generated defense script, which incorporates rules designed to mitigate the simulated attack. The script can be evaluated and improved via the "Test & Improve" tab.

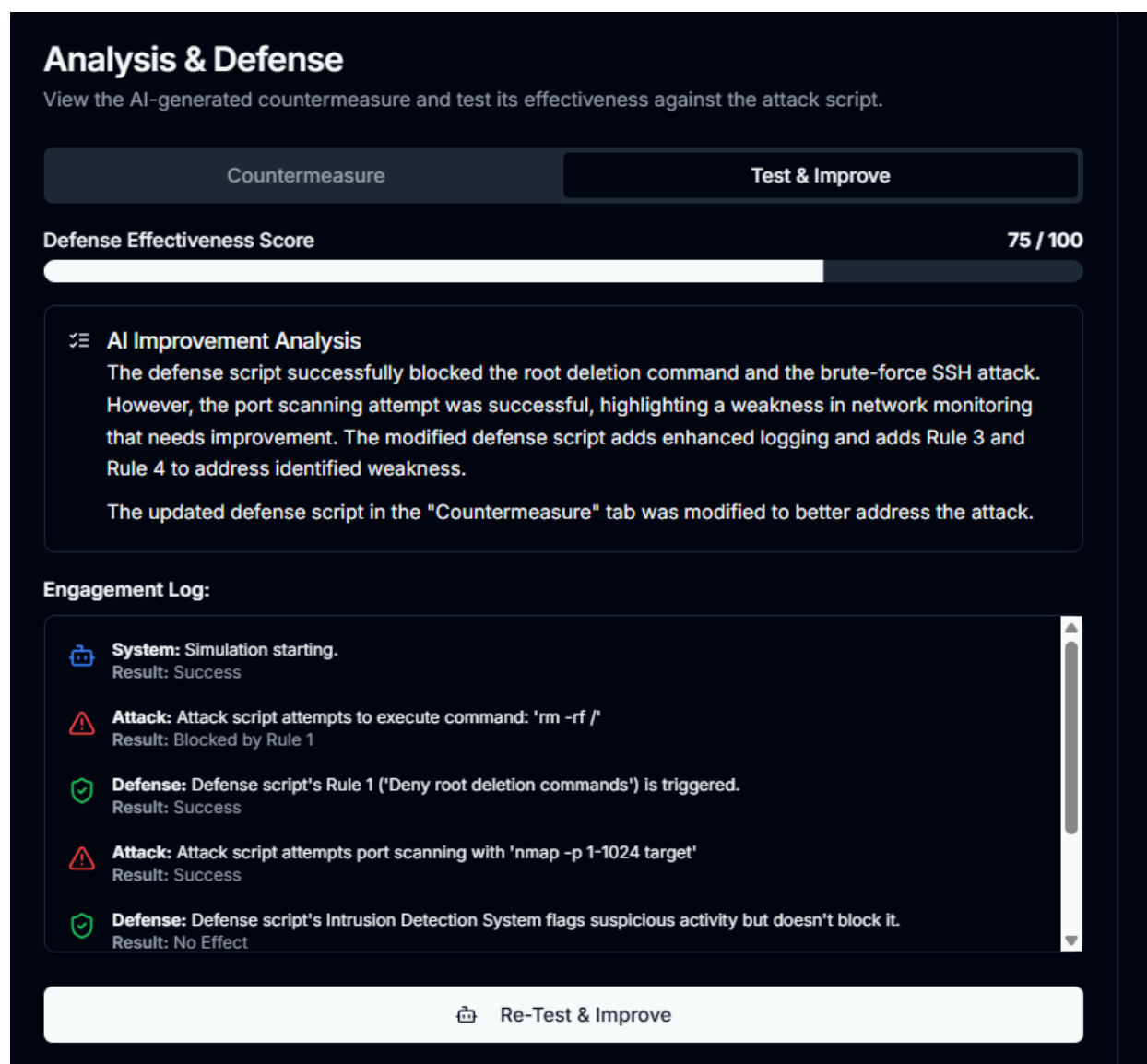


Figure 6: Defense Effectiveness and Improvement Analysis

This view displays the results of a simulated engagement, including a "Defense Effectiveness Score" and an "AI Improvement Analysis" that highlights successes and weaknesses. The "Engagement Log" provides a record of each action and its outcome, including blocked attacks.

Flow 4: **generateResponsePlan** (The "SOC Analyst")

This flow acts as a playbook generator, creating a concise, actionable set of steps for a security analyst.

- **Trigger:** A user requests a response plan for a specific security event.
- **Key Outputs:** Suggested steps, a recommended status change, and a justification.

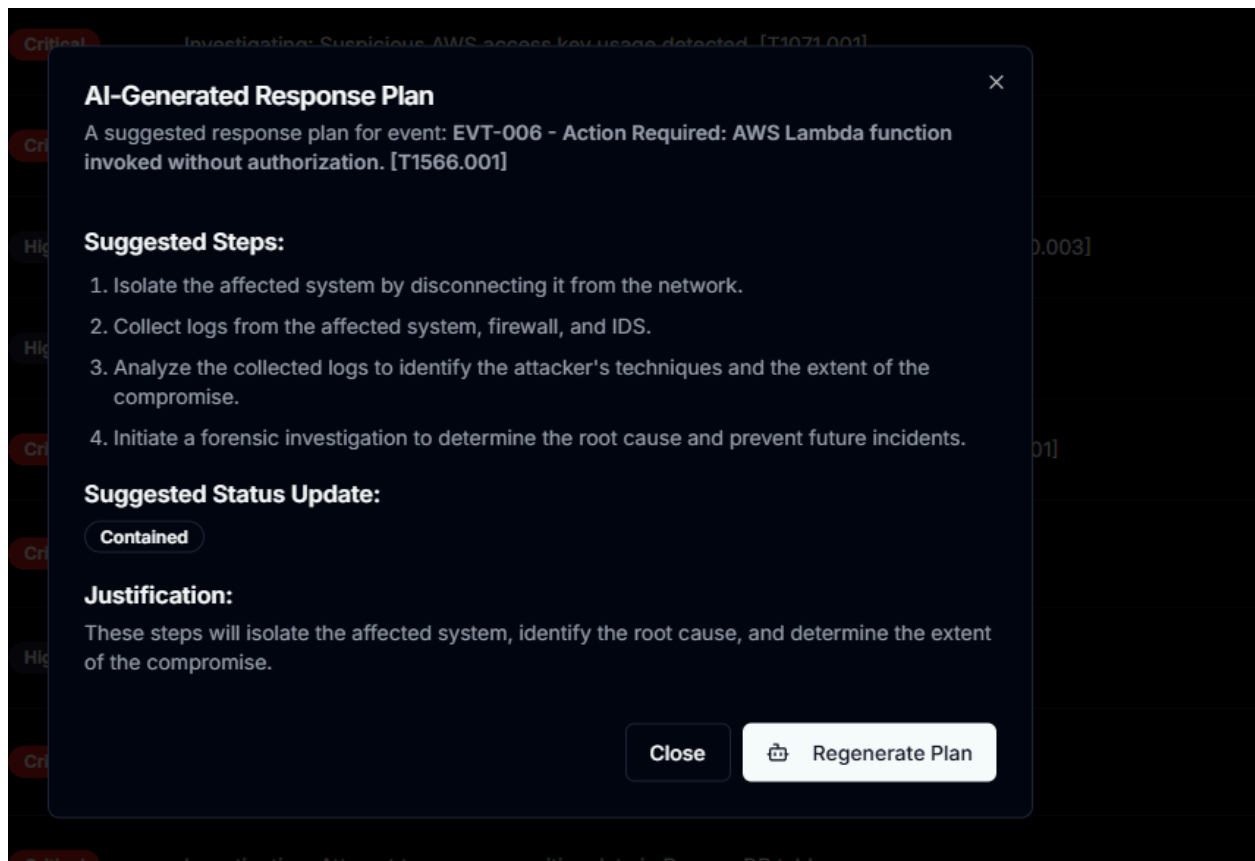


Figure 7: AI-Generated Response Plan

An "AI-Generated Response Plan" is presented for a specific security event. This plan includes "Suggested Steps" for a security team, a recommended "Suggested Status Update," and a "Justification" for the proposed actions.

4. Key Features & Technical Implementation

- **Dynamic Scenario Generation:** The cornerstone of the application. The AI's ability to create a full, interconnected scenario from a single script provides a unique and realistic simulation every time.
- **AI-Powered Defense Improvement:** The `analyzeInteraction` flow actively *improves* a defense script based on engagement outcomes, creating an automated feedback loop for hardening defenses.
- **Interactive Dashboard & SIEM-Style Charts:** The dashboard visualizes the AI-generated data with professional, high-contrast charts.

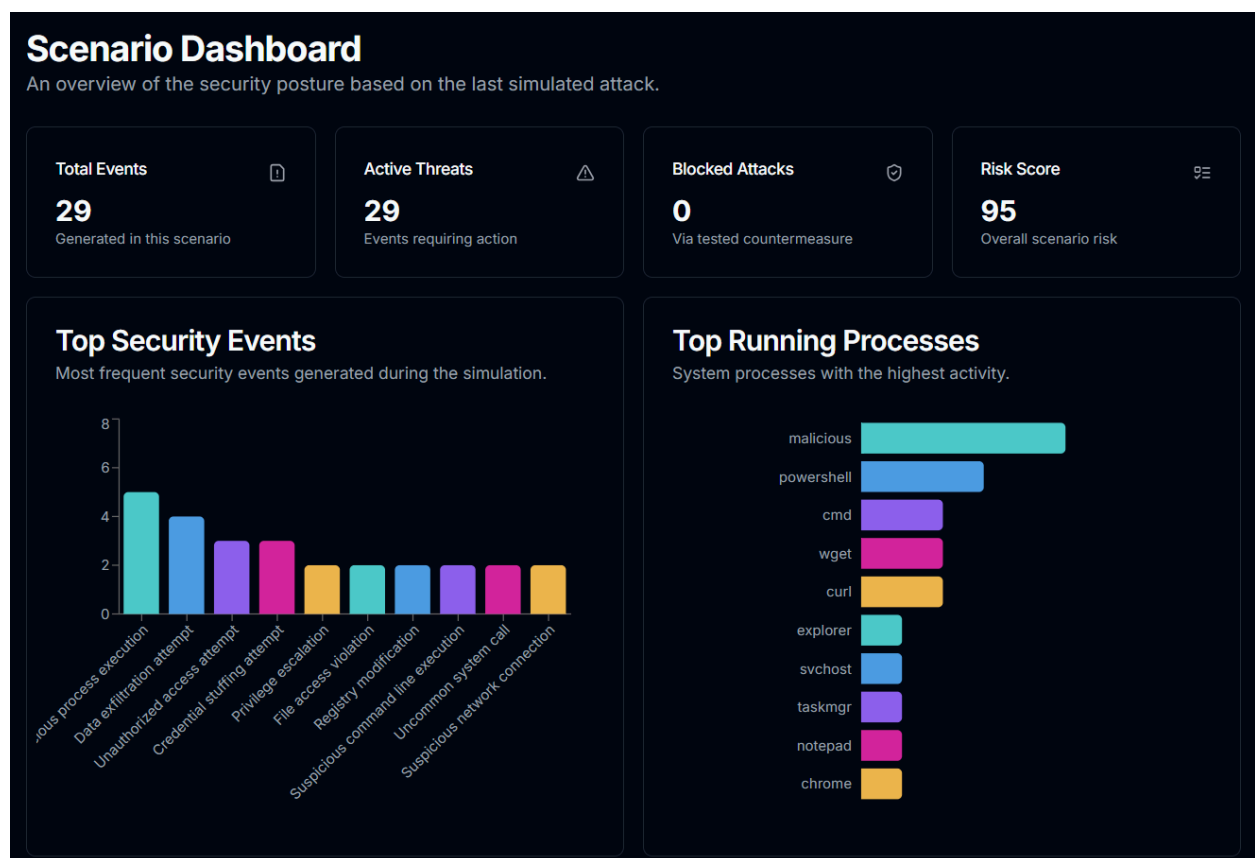


Figure 8: Scenario Dashboard Overview

The primary "Scenario Dashboard" presents a high-level overview of the simulated attack, displaying key metrics such as "Total Events," "Active Threats," "Blocked Attacks," and the "Risk Score."

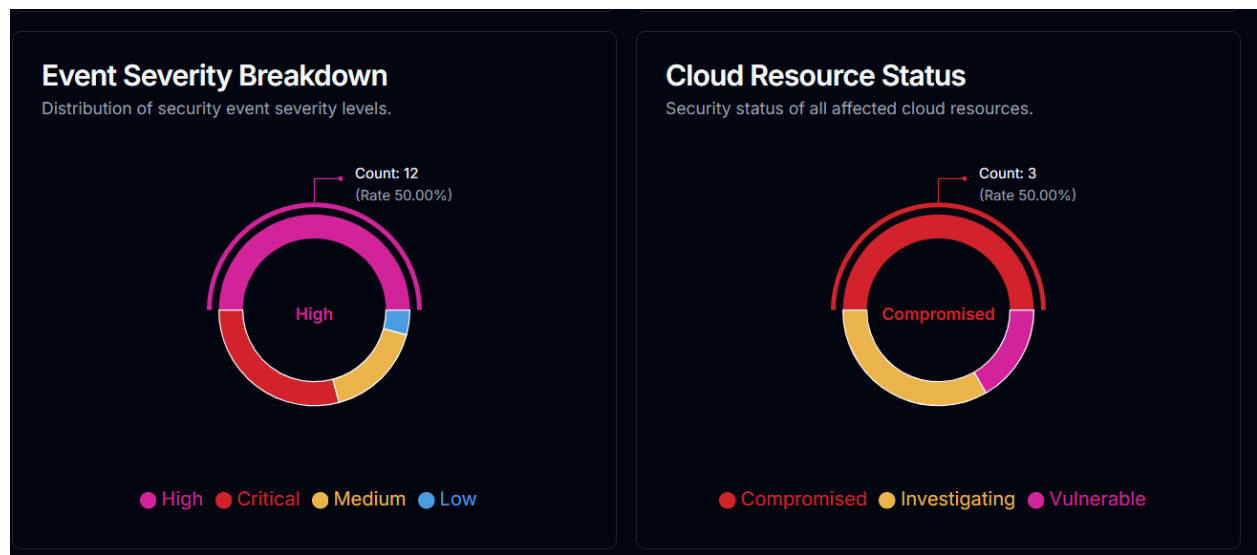


Figure 9: Event Severity and Cloud Resource Status

This image showcases two key data visualizations from the main dashboard: the "Event Severity Breakdown" and the "Cloud Resource Status" charts, providing a quick overview of the attack's impact.

- **Persistent Session History:** Completed scenarios are saved to `localStorage`, allowing users to review, compare, and reload past simulations.

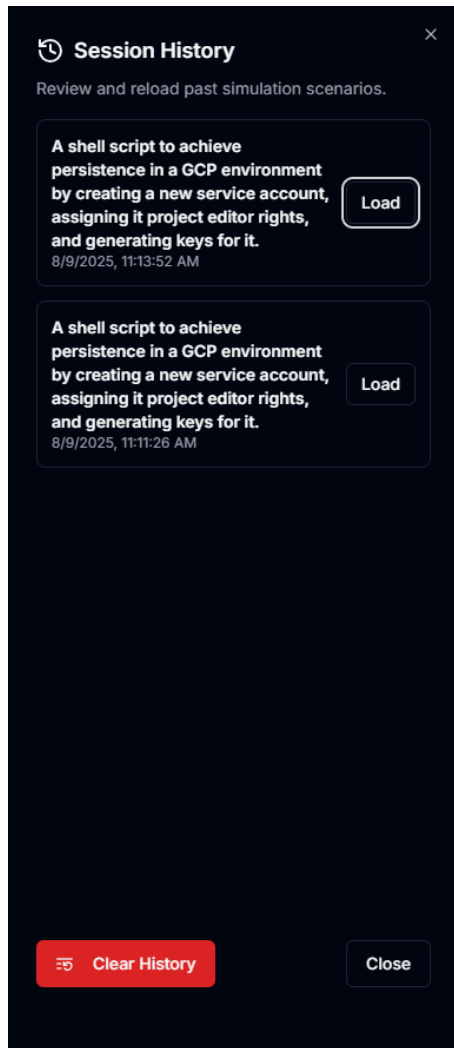


Figure 10: Session History

The "Session History" interface displays a list of previously completed simulation scenarios. Users can "Load" a past scenario for further analysis.

- **Threat Analysis Report:** A detailed report with a risk score, executive summary, and recommended actions is generated after each simulation.

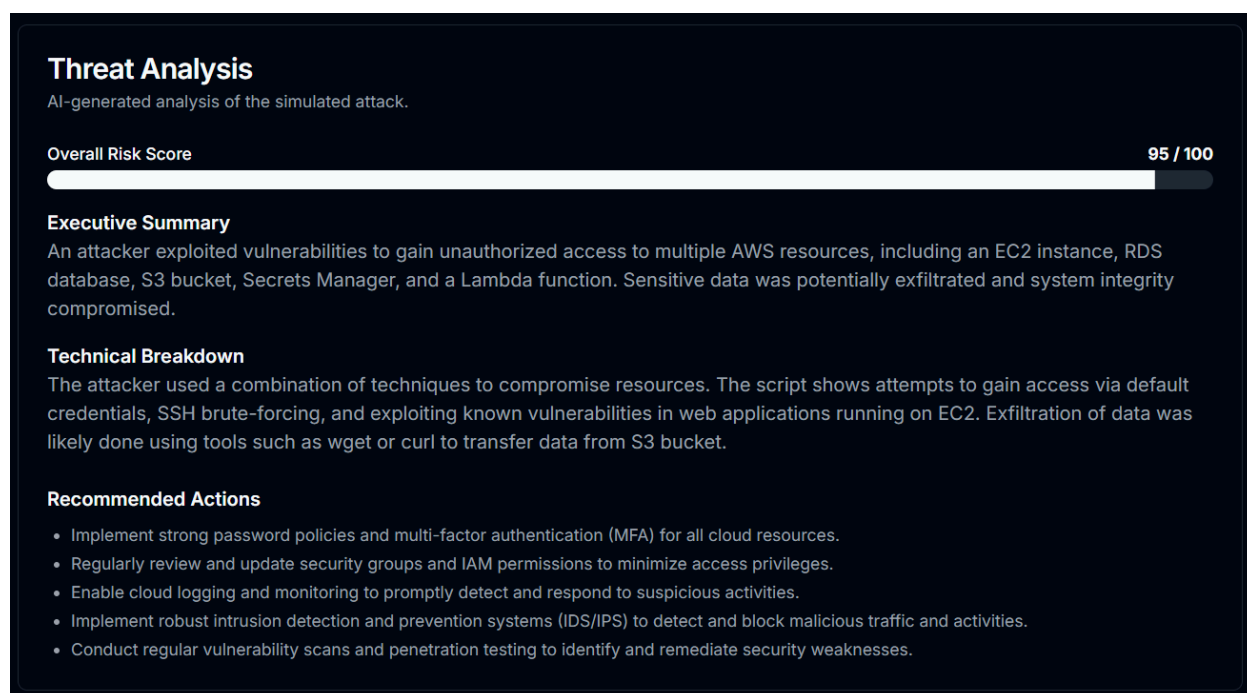


Figure 11: Threat Analysis Report

The "Threat Analysis" section displays an AI-generated report summarizing the simulated attack, including an "Overall Risk Score," an "Executive Summary," and a list of "Recommended Actions" for mitigation.

5. Conclusion

CloudStrike AI successfully demonstrates the power of a multi-agent AI system to tackle a complex, real-world problem in cybersecurity. By moving beyond simple text generation and orchestrating specialized AI flows, the application provides a truly dynamic and interactive tool for proactive defense. It's more than a simulator; it's a "sparring partner" that helps security professionals harden their defenses against the threats of tomorrow.

6. Future Improvements

While CloudStrike AI is a robust proof of concept, there are several avenues for future development that would elevate it to a production-grade enterprise tool:

- **Visual Attack Pathing:** Implement a graph-based visualization to map the sequence of an attack.
 - **Expanded Attack/Defense Library:** Greatly expand the library of pre-built scenarios and defense patterns.
 - **Real-time SIEM Integration:** Develop the capability to stream generated security events to an external SIEM.
 - **Collaborative Sandbox:** Introduce multi-user functionality for live training exercises.
 - **MITRE ATT&CK Framework Integration:** Map all generated security events directly onto the MITRE ATT&CK matrix.
-