PROJECT ABSTRACT

Y2K-Cyber-AI: An Agentic Dual-Mode Cyber Defense and Offensive Simulation Platform

Y2K-Cyber-AI is an advanced AI-driven cybersecurity platform designed to integrate defensive security operations and controlled offensive simulation within a unified agentic framework. The system operates through a single intelligent AI agent that dynamically switches between Blue Mode (Defensive Intelligence) and Red Mode (Offensive Simulation), enabling real-time threat detection, malware analysis, automated response, and secure adversarial testing within authorized lab environments.

In Blue Mode, the platform functions as an AI-powered Security Operations Center (SOC). It performs real-time log ingestion, behavioral anomaly detection, dynamic malware analysis using isolated sandbox environments, risk scoring, and automated incident response. The system integrates external threat intelligence APIs such as VirusTotal and leverages AI reasoning models to correlate behavior patterns with known attack techniques. It provides mitigation guidance, compliance insights, and defensive playbook generation to strengthen organizational resilience.

In Red Mode, Y2K-Cyber-AI acts as a controlled offensive simulation assistant designed for ethical testing within user-owned virtual machines. Instead of executing exploits on external systems, all tools and commands are executed securely through SSH on user-managed lab VMs. The AI designs attack paths conceptually, simulates adversarial strategies, evaluates vulnerabilities, and teaches secure system design principles. This ensures responsible cybersecurity education and safe red-team training.

A key innovation of the platform is its dynamic malware analysis framework. Suspicious files are executed exclusively within isolated sandbox virtual machines created through snapshot-based environments. The AI monitors process behavior, registry modifications, file system changes, and network activity, generating comprehensive behavioral reports and defensive recommendations.

By combining AI-driven automation, sandbox-based dynamic analysis, SSH-restricted tool execution, and dual-mode operational intelligence, Y2K-Cyber-AI transforms traditional cybersecurity tools into an interactive AI cyber range platform. The system serves as a bridge between detection, defense, adversarial simulation, and cybersecurity education, providing a scalable and secure approach to modern threat management.

ARCHITECTURE EXPLANATION SECTION

1️ Overall Architectural Vision

Y2K-Cyber-AI follows a **Layered Agentic Architecture** built around a single AI core that dynamically adapts its behavior based on operational mode.

The system is divided into five major layers:

Presentation Layer

Application Layer

Agent Intelligence Layer

Execution & Sandbox Layer

External Integration Layer

2️⃣ Presentation Layer (Frontend)

Technology:

- React.js
- Tailwind CSS
- Real-time WebSocket integration

Features:

- Mode Toggle (Blue / Red)
- Malware upload interface
- Log monitoring dashboard
- Attack simulation visualizer
- VM connection management (SSH input)
- API key configuration panel

The UI dynamically changes color theme:

- 🔵 Blue Mode – Defensive dashboard
- 🔴 Red Mode – Offensive simulation dashboard

3️⃣ Application Layer (Backend Core)

Technology:

- Node.js (Express)
- FastAPI (AI Engine Microservice)

Responsibilities:

- Authentication & RBAC
- Mode switching
- API key handling
- Log storage
- Session management
- SSH orchestration
- Tool request validation

This layer ensures:

- No direct execution on platform server
- Strict sandbox enforcement
- Secure API handling

4️⃣ Agent Intelligence Layer (Core Brain)

This is the heart of Y2K-Cyber-AI.

🧠 Single Agent – Dual Personality Model

Y2K-Agent Core

├— Blue Personality Logic

├— Red Personality Logic

├— Supervisor Controller

├— Memory System (Vector DB)

└— Tool Decision Engine

Blue Mode Behavior:

- Analyzes logs
- Detects anomalies
- Correlates with threat intelligence
- Generates defensive recommendations
- Triggers automated response

Red Mode Behavior:

- Designs simulated attack flows
- Performs vulnerability reasoning
- Executes tools only on user VM
- Teaches secure system design

Supervisor Controller:

- Verifies tool permissions
- Restricts execution to authorized VM
- Prevents external scanning
- Logs all AI decisions

5️⃣ Execution & Sandbox Layer

This is the most critical security component.

🧪 A. Dynamic Malware Analysis Sandbox

Architecture:

Host System

└— Virtual Machine (Snapshot Enabled)

├— Internal Network Only

├— Monitoring Agents (Sysmon)

├── Process Monitor

├── File Change Tracker

└── Network Traffic Capture

Workflow:

1. User uploads suspicious file
2. File transferred to isolated VM
3. VM snapshot restored
4. File executed
5. Behavior monitored
6. Logs collected
7. VM reset to clean snapshot

This ensures:

- No host infection
- No persistent compromise
- Controlled behavioral analysis

🖥️ B. SSH-Based Tool Execution Model

For ethical control:

- User provides SSH credentials to their own lab VM
- AI sends tool commands via SSH
- Results streamed back
- No tool runs on Y2K server

Flow:

User Request

→ AI Agent

→ Supervisor Validation

→ SSH Command to User VM

→ Output Returned

→ AI Analysis

This ensures:

- Platform cannot be misused
- All offensive testing remains local
- Full accountability

6️⃣ External Integration Layer

Y2K-Cyber-AI integrates multiple APIs:

- VirusTotal → File reputation
- Gemini/OpenAI → AI reasoning
- AbuseIPDB → IP reputation
- Shodan (optional) → Infrastructure insights

API keys:

- Provided by user
- Stored encrypted
- Never exposed publicly

7️⃣ Data Flow Overview

Blue Mode Data Flow:

Logs → AI Analysis → Threat Correlation

→ Sandbox (if needed)

→ Risk Score

→ Automated Action

→ Report Generation

Red Mode Data Flow:

User Request → AI Planning

→ SSH Tool Execution

→ Result Analysis

→ Attack Simulation Model

→ Defensive Mapping (Blue Comparison)

8️⃣ Security & Ethical Control Framework

Y2K-Cyber-AI enforces:

- VM-only execution
- Snapshot restoration
- No external IP scanning
- Mode-based restrictions
- Tool usage logging
- Supervisor validation engine

This makes it:
✓ Ethical
✓ Educational
✓ Controlled
✓ Enterprise-ready

🧠 Key Innovations in Architecture

1. Single Agent – Dual Operational Personalities
2. Sandbox-Driven Dynamic Malware Analysis
3. SSH-Restricted Tool Execution Model
4. AI-Guided Defensive and Offensive Learning
5. Integrated Cyber Range Functionality

🔥 Final Positioning

Y2K-Cyber-AI is not just:

- A SIEM
- A Red Team tool
- A Malware sandbox

It is: An Agentic AI Cyber Range Platform

That Thinks, Teaches, Defends, and Simulates.