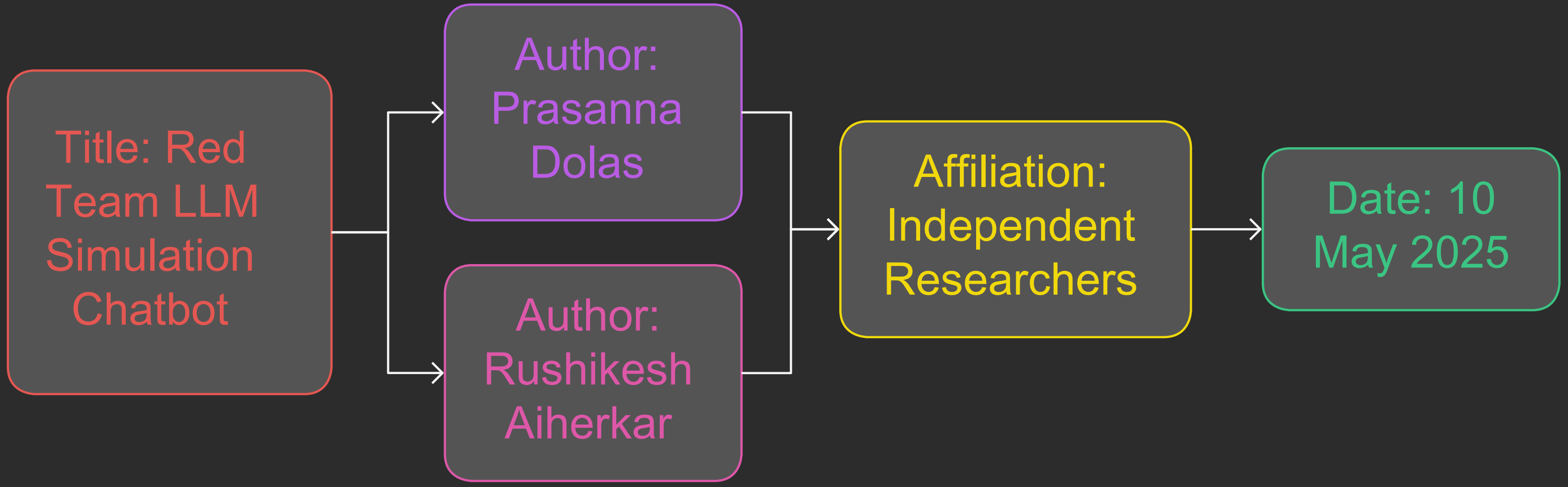
LLM-based Red Team Simulations

Red Team LLM Simulation Chatbot Development



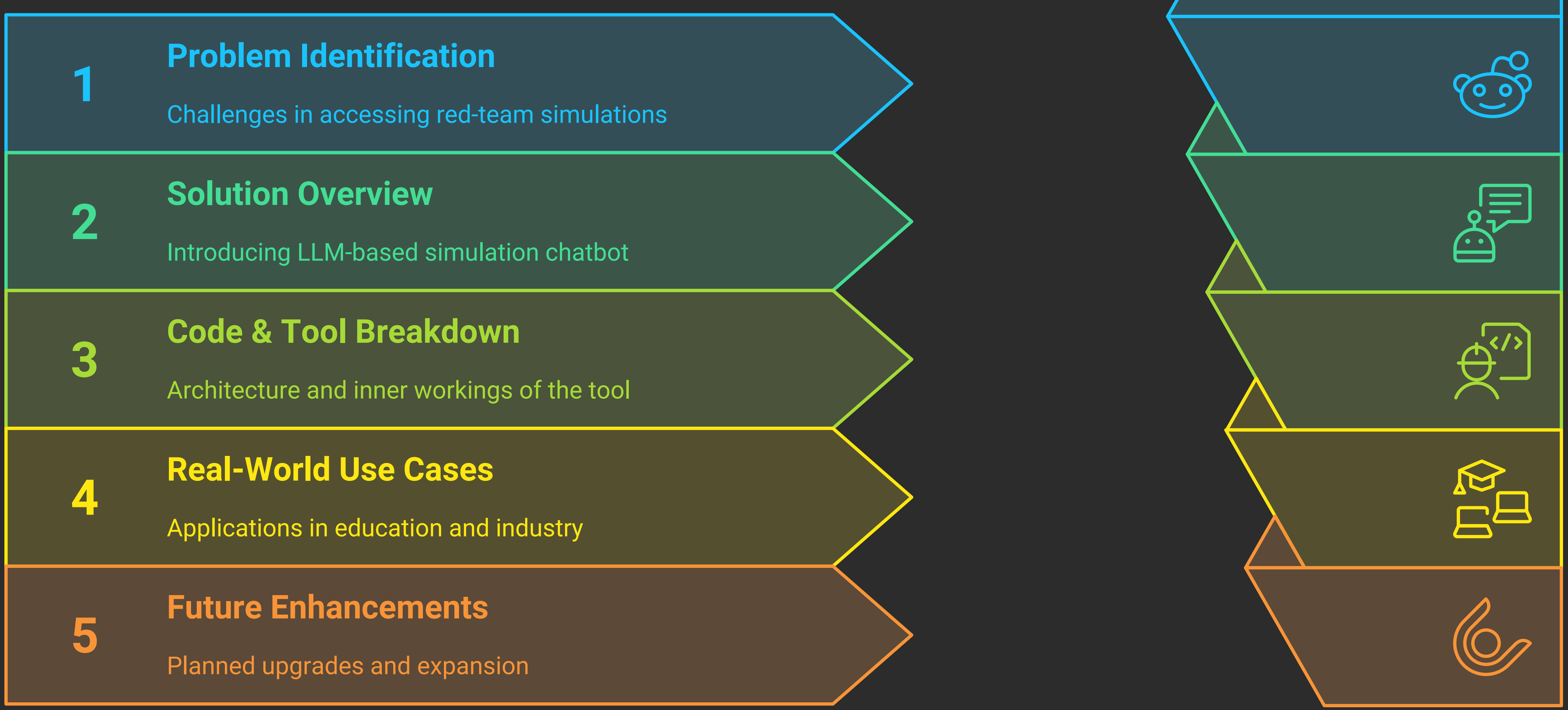
Title: Red Team LLM Simulation Chatbot

Author: Prasanna Dolas

Author: Rushikesh Aiherkar

Affiliation: Independent Researchers

Date: 10 May 2025



1. **Problem Identification**
   Challenges in accessing red-team simulations

2. **Solution Overview**
   Introducing LLM-based simulation chatbot

3. **Code & Tool Breakdown**
   Architecture and inner workings of the tool

4. **Real-World Use Cases**
   Applications in education and industry

5. **Future Enhancements**
   Planned upgrades and expansion

LLM-Driven Cybersecurity Training Cycle

1

Identify Threats

Recognize evolving cybersecurity threats

2

Conduct Red-Teaming

Simulate attacks to test defenses

3

Evaluate Resources

Assess the cost and availability of resources

4

Implement LLM Solution

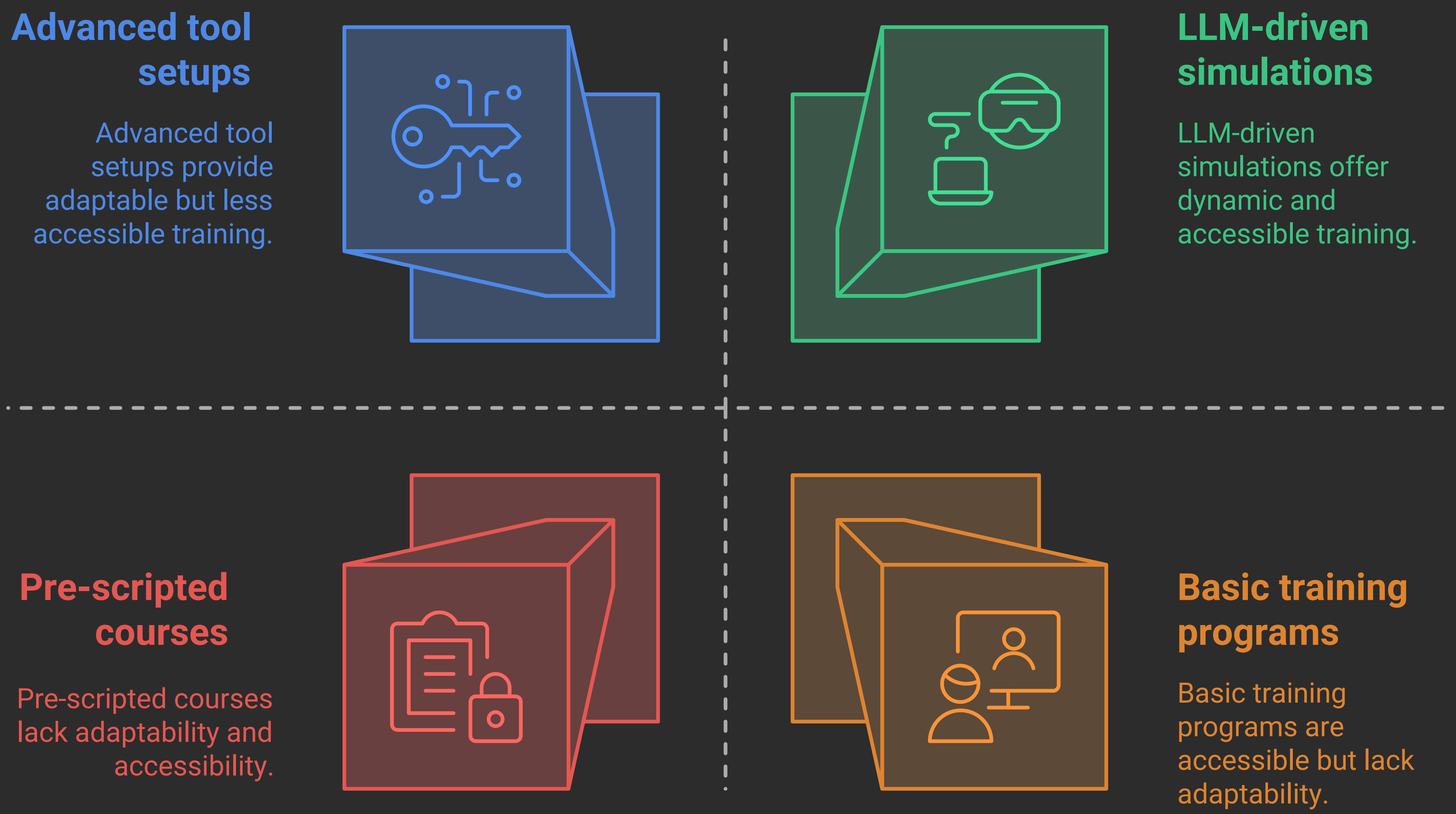
Use LLMs to create cost-effective training

5

Train Defenders

Educate cybersecurity professionals

Challenges in Red-Team Training



Advanced tool setups

Advanced tool setups provide adaptable but less accessible training.

LLM-driven simulations

LLM-driven simulations offer dynamic and accessible training.

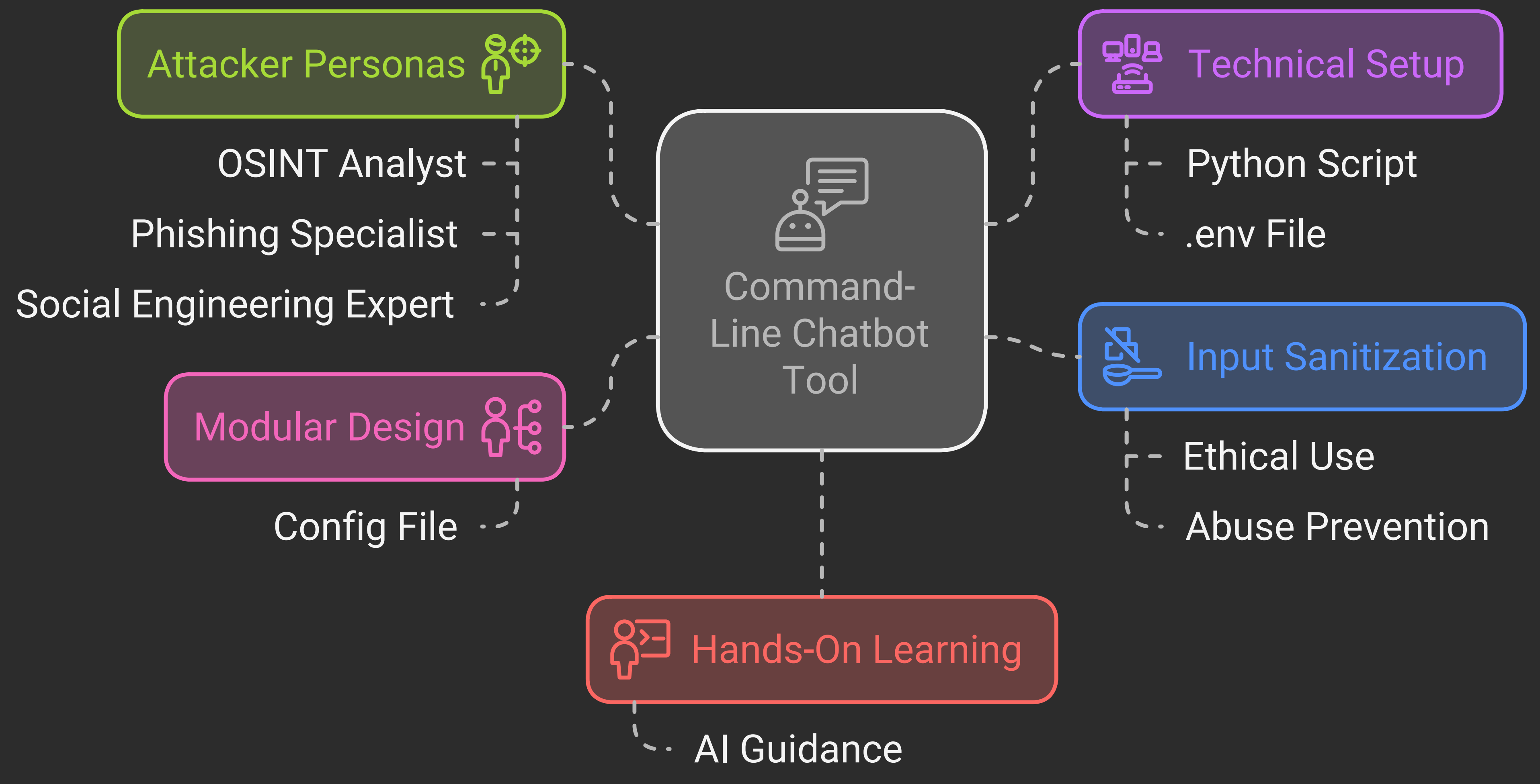
Pre-scripted courses

Pre-scripted courses lack adaptability and accessibility.

Basic training programs

Basic training programs are accessible but lack adaptability.

Solution Overview: Command-Line Chatbot Tool



Command-Line Chatbot Tool

Attacker Personas

- OSINT Analyst
- Phishing Specialist
- Social Engineering Expert

Modular Design

- Config File

Technical Setup

- Python Script
- .env File

Input Sanitization

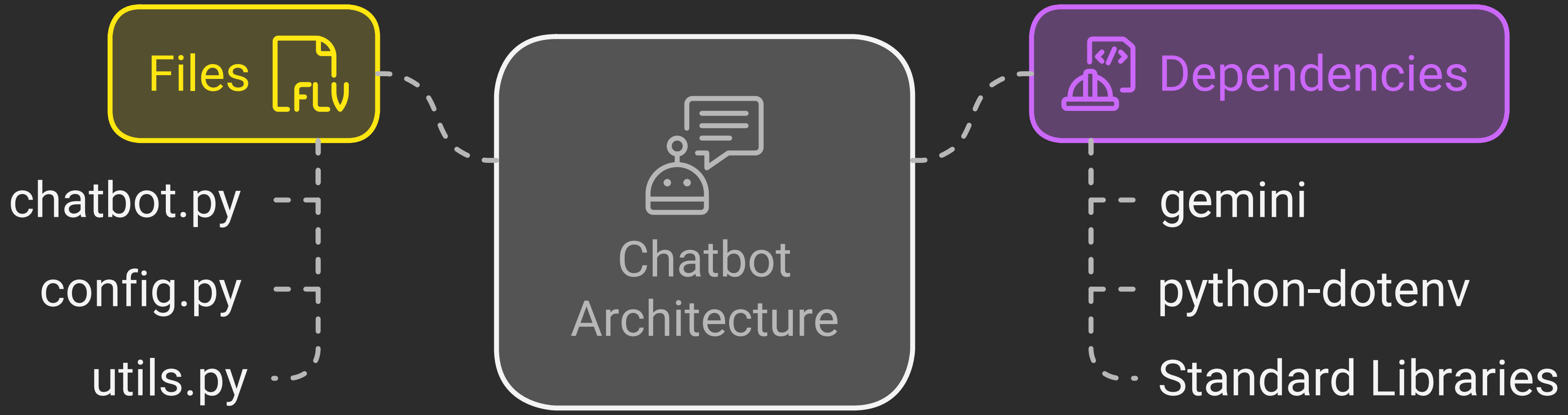
- Ethical Use
- Abuse Prevention

Hands-On Learning

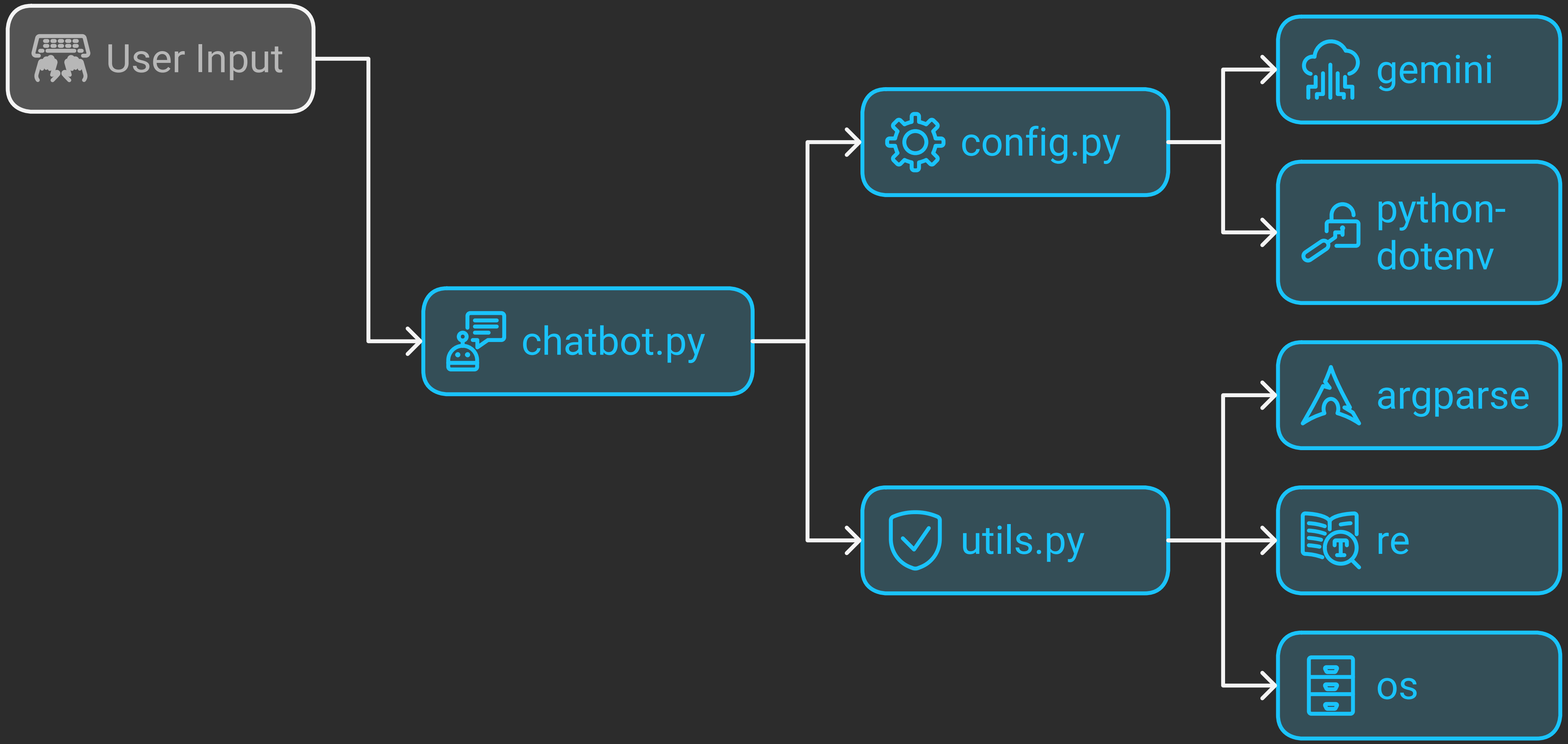
- AI Guidance

Made with Napkin

Chatbot Architecture and Dependencies

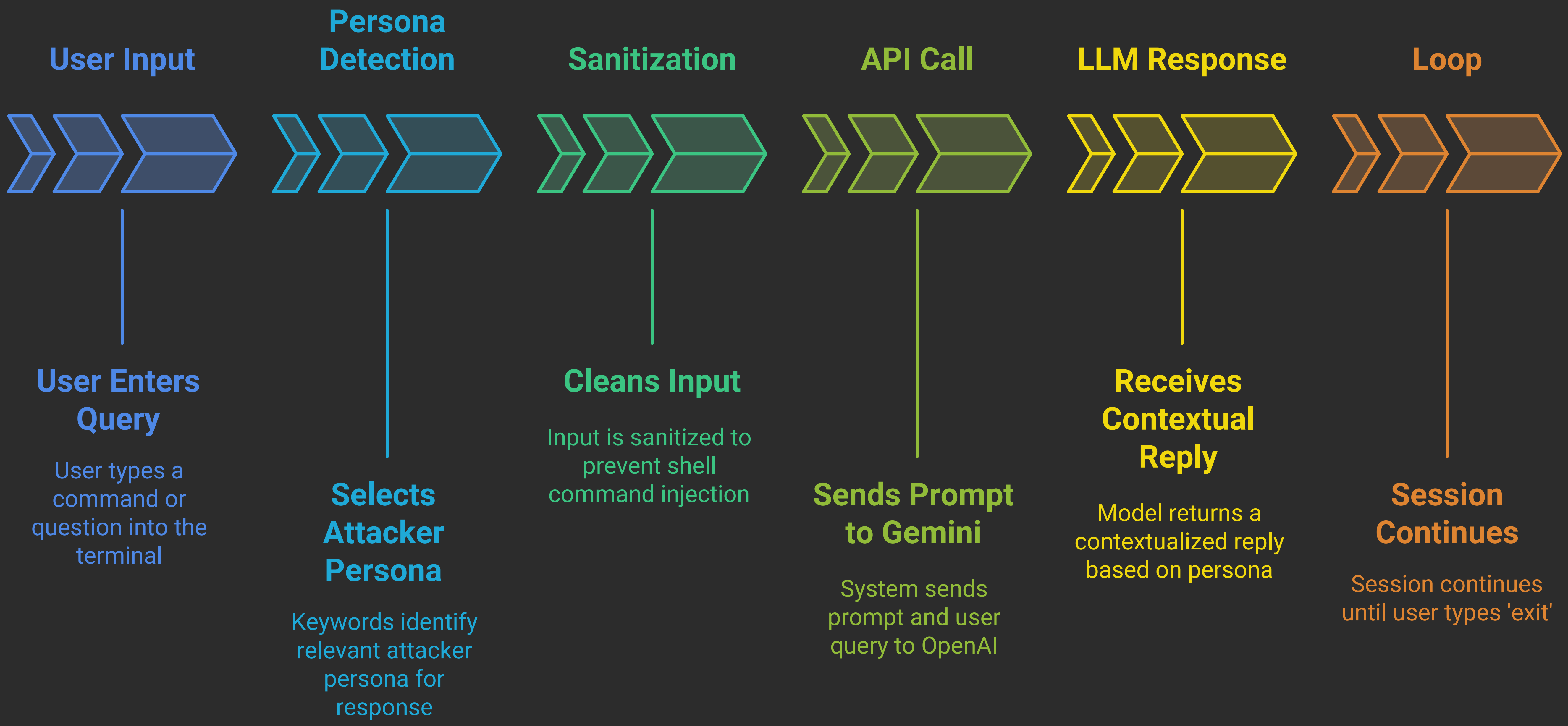


Chatbot System Architecture



Made with Napkin

Chatbot Workflow Steps



User Input

User Enters Query

User types a command or question into the terminal

Persona Detection

Selects Attacker Persona

Keywords identify relevant attacker persona for response

Sanitization

Cleans Input

Input is sanitized to prevent shell command injection

API Call

Sends Prompt to Gemini

System sends prompt and user query to OpenAI

LLM Response

Receives Contextual Reply

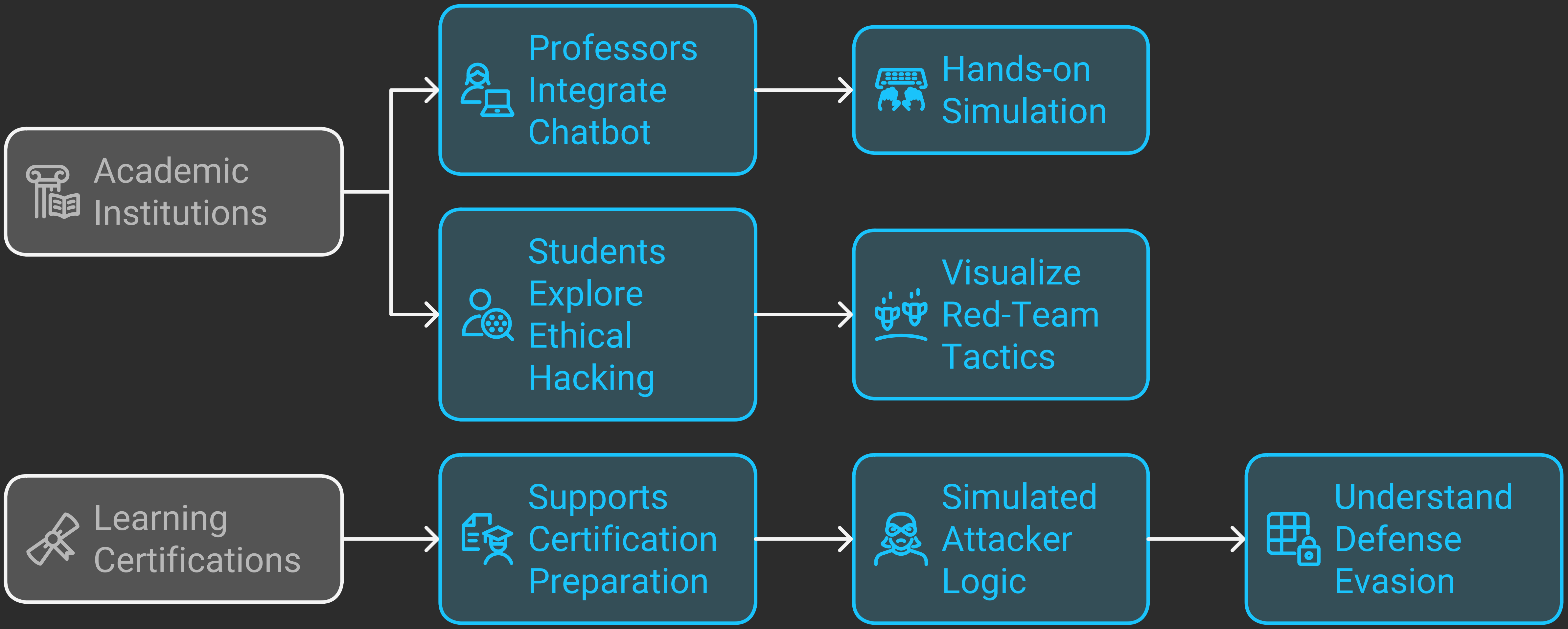
Model returns a contextualized reply based on persona

Loop

Session Continues

Session continues until user types 'exit'

Cybersecurity Chatbot Integration in Education



Academic Institutions

Professors Integrate Chatbot

Hands-on Simulation

Students Explore Ethical Hacking

Visualize Red-Team Tactics

Learning Certifications

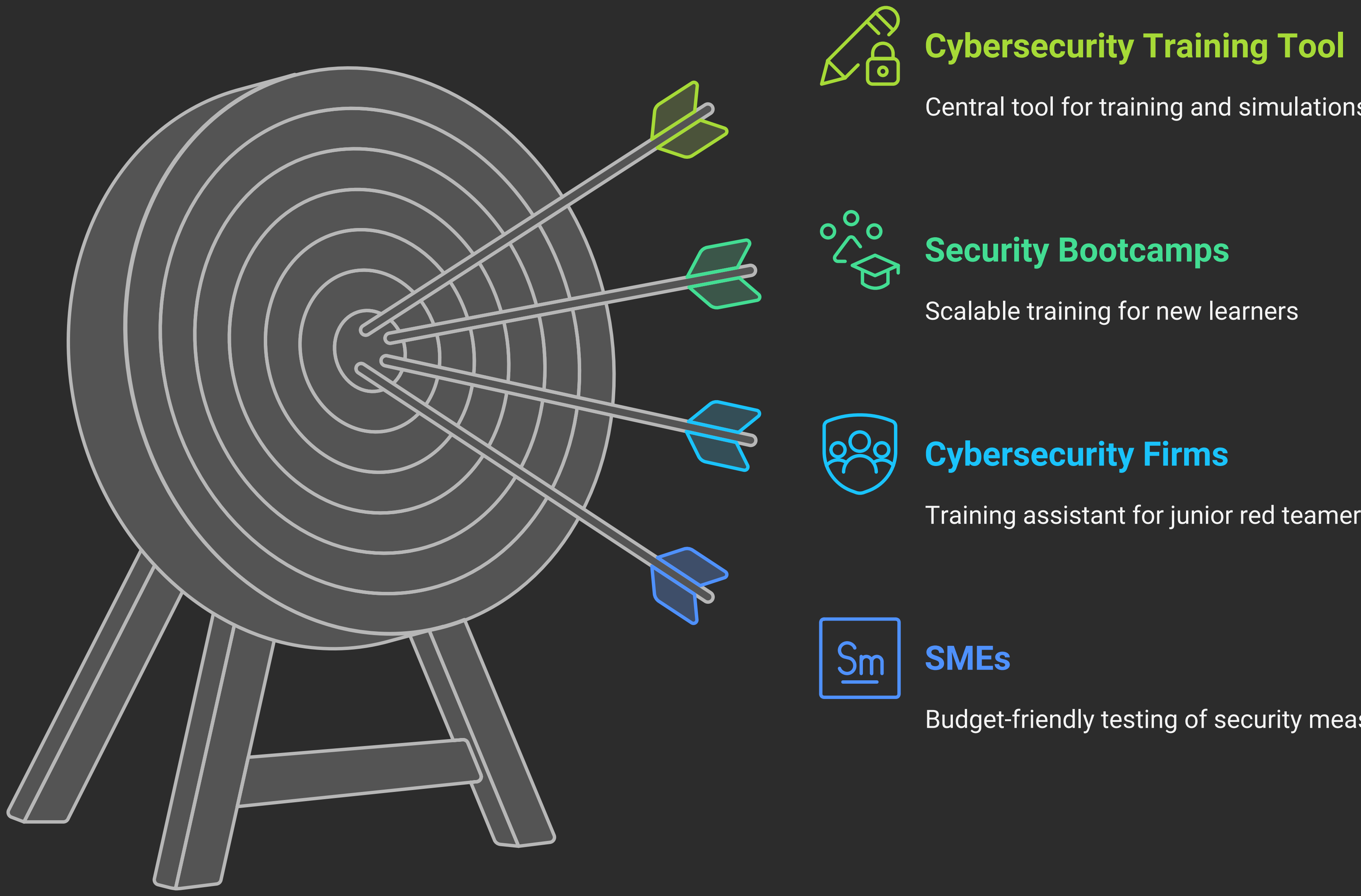
Supports Certification Preparation

Simulated Attacker Logic

Understand Defense Evasion

Made with Napkin

Tool Application for Cybersecurity Training



Cybersecurity Training Tool

Central tool for training and simulations

Security Bootcamps

Scalable training for new learners

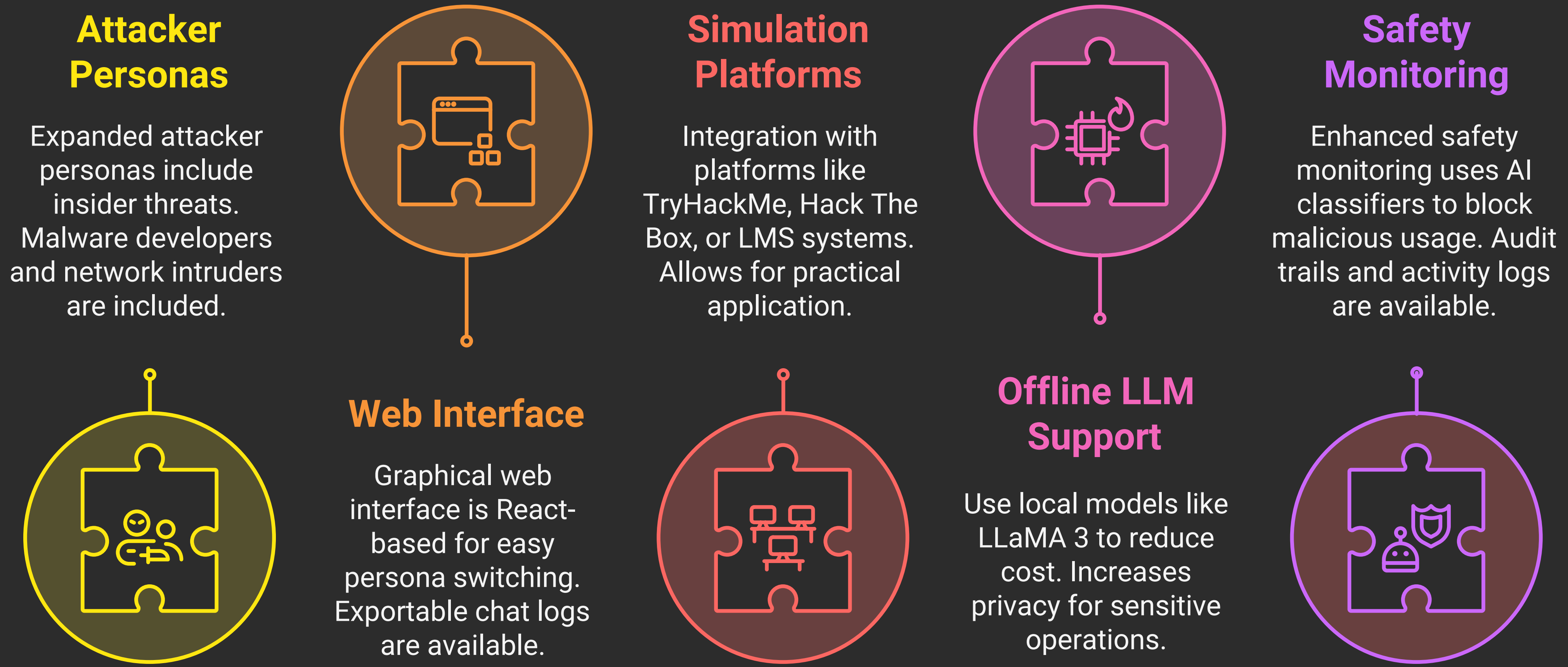
Cybersecurity Firms

Training assistant for junior red teamers

SMEs

Budget-friendly testing of security measures

Software Features



Attacker Personas

Expanded attacker personas include insider threats. Malware developers and network intruders are included.

Web Interface

Graphical web interface is React-based for easy persona switching. Exportable chat logs are available.

Simulation Platforms

Integration with platforms like TryHackMe, Hack The Box, or LMS systems. Allows for practical application.

Offline LLM Support

Use local models like LLaMA 3 to reduce cost. Increases privacy for sensitive operations.

Safety Monitoring

Enhanced safety monitoring uses AI classifiers to block malicious usage. Audit trails and activity logs are available.