Empowering a Breach & Attack Simulation (BAS) Platform using a Large Language Model (LLM) to improve Security Tests

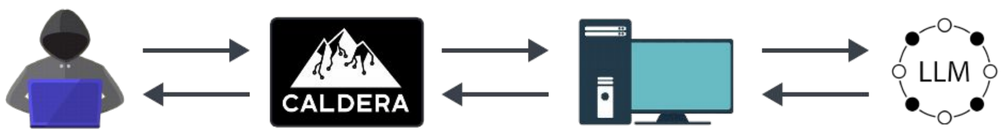
Abstract

This project presents an architecture integrating a Large Language Model (LLM) and a Breach & Attack Simulation platform (BAS); tested with MITRE Caldera. Designed as an intelligent copilot for penetration testing, this copilot recommends Operating Systems commands to identify and validate vulnerabilities in a simulated environment, enhancing adaptivity and interactivity for cybersecurity professionals.

Objectives

Objectives:

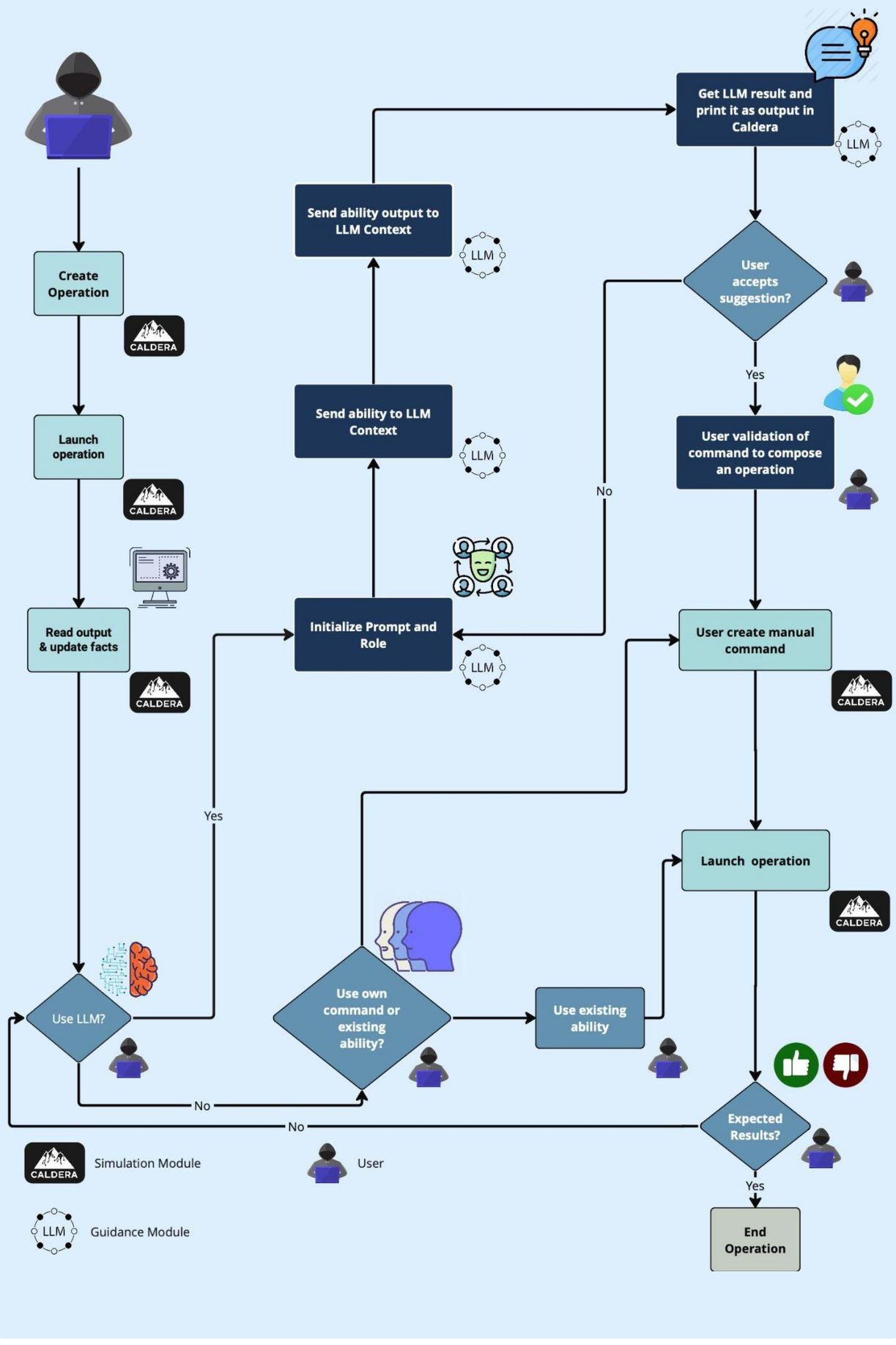
- **Guided Vulnerability Detection:** Leverage LLM-driven insights to identify and address vulnerabilities effectively.
- **Educational Support:** Enable cybersecurity professionals to better understand attack scenarios through context-specific LLM recommendations.



Background

Small and medium-sized enterprises face increasing cybersecurity threats but lack resources for effective defense, as traditional pentesting is costly and complex, exposing them to significant financial and operational risks. The solution builds on pentesting, BAS tools (like MITRE Caldera), and advancements in AI (LLMs like GPT) to make these technologies more accessible.

Proposal



Reached Results

Exhaustive and complete experiments conducted on a Metasploitable 2-3 Virtual Machine. The following vulnerabilities identified and successfully exploited: **Backdoor**, **Web Server Shutdown** and **Privilege Escalation**.

4 Use Cases

1. BAS operates automatically
2. BAS starts operation, then the control is taken by the LLM to finish the operation
3. Control alternates between the BAS and LLM
4. LLM performs in the operation by itself

Conclusions

Our project introduced an LLM-BAS copilot for penetration testing by integrating an LLM with a BAS platform, which is a combination not yet available on the market. This innovative tool guides users in real-time, automates key steps, and enhances vulnerability detection, making cybersecurity testing faster, smarter, and more accessible.

References

- Martínez Sánchez et al. (2024). Methodology for Automating Attacking Agents in Cyber Range Training Platforms. University of Murcia, Spain; SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France.
- MITRE.(n.d.).Caldera. https://caldera.mitre.org
- IBM. (2023). Cost of a data breach 2023. https://www.ibm.com/reports/data-breach
- Deng, G., Liu, Y., Mayoral-Vilches, V., Liu, P., Li, Y., Xu, Y., Zhang, T., et al. (2024). PentestGPT: An LLM-empowered automatic penetration testing tool. arXiv. https://arxiv.org/abs/2308.06782

Autor(es)

Sebastián Caballero-Roa, Juan Martínez-Páez, Juan Velandia-Romero, Daniel Díaz-López
Matemáticas Aplicadas y Ciencias de la Computación (MACC)
Escuela de Ingeniería, Ciencia y Tecnología (EICT)
Universidad del Rosario

Organiza

Universidad Nacional de Colombia
Grupo de investigación UNsecureLab
Semillero de investigación Uqbar
Grupo de investigacion Tlön

UNIVERSIDAD NACIONAL DE COLOMBIA