



UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II

itee_{PhD}
information technology
electrical engineering



DIE
TI

UNI
NA

Luciano Pianese

Automated Offensive Security in the Age of Generative AI

Tutor: Roberto Natella

co-Tutor: Marco Braccioli

Cycle: XXXIX

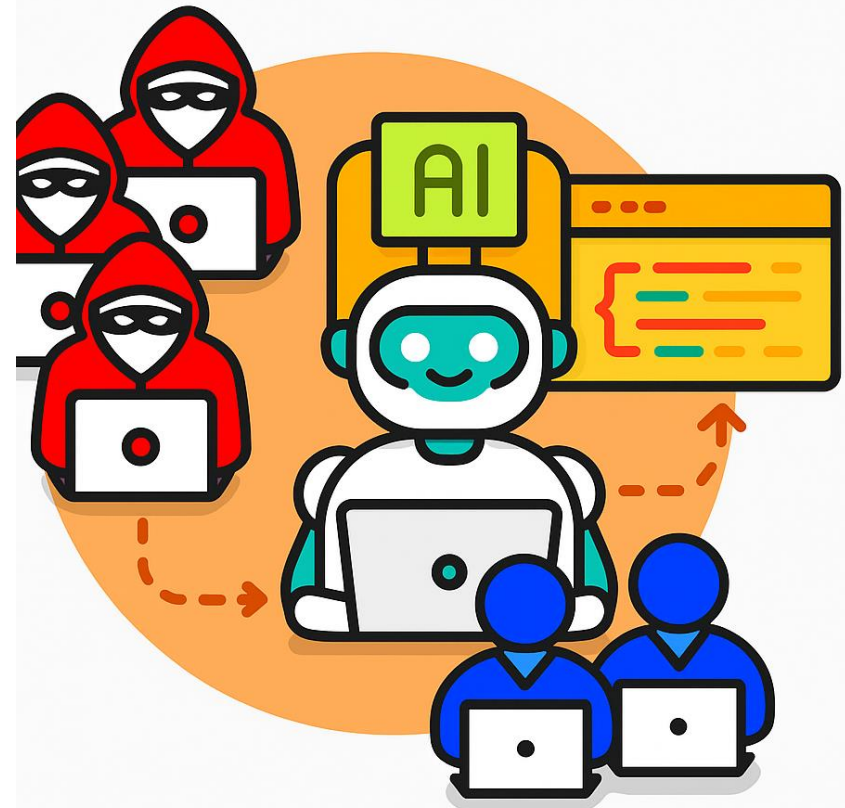
Year: Second

Background

- **M.Sc. Degree** in Computer Engineering –
Università degli Studi di Napoli Federico II
- **Research Group:** DESSERT
- **Ph.D. start date:** 01/11/2023
- **Scholarship type:** PNRR - DM 117/2023, Missione 4
- Componente 2 - Investimento 3.3 Dottorati innovative
- **Partner company:** DigitalPlatforms S.p.A.

Research field of interest

Focused on the intersection of **Offensive Security** and **Generative Artificial Intelligence (AI)**, aiming to develop **methodologies for creating and evaluating autonomous systems** capable of automating offensive security procedures



Summary of study activities

	Courses	Seminars	Research	Tutorship
Total	32	10.2	76.8	1
Expected	30-60	10-20	40-80	0-3.2

- **PhD courses:**
 - Strategic Orientation for STEM Research & Writing
 - Hands-on Network Intrusion Detection via Machine and Deep Learning
 - Virtualization technologies and their applications
 - Statistical data analysis for science and engineering research
 - Using Deep Learning properly
 - Innovation and Entrepreneurship
 - IELTS Advanced preparation course

Research Activity (1), problem statement

- Problem:
 - The rapid advancement of Large Language Models (LLMs) has significantly enhanced the capabilities available to attackers
- Objectives:
 - A methodology to generate and analyze complete cyber attacks in post-compromised scenario

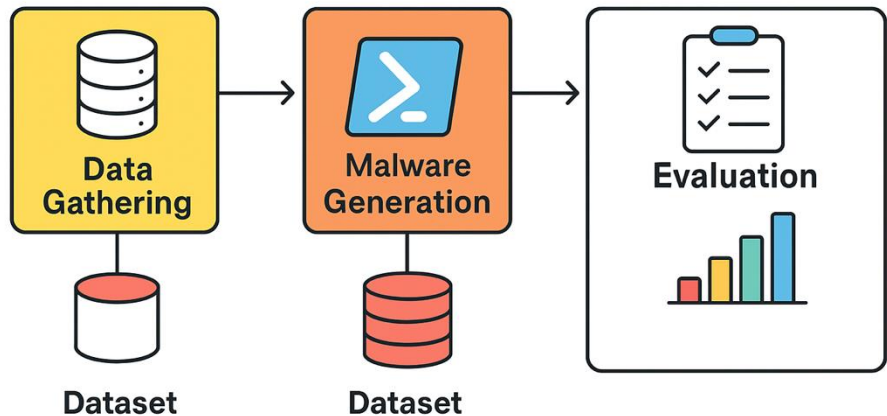
The collage features several documents from ENISA:

- ENISA 20th Anniversary Banner:** Includes the ENISA logo, the text "enisa 20 years!", and the European Union Agency for Cybersecurity logo.
- Report: "Vibe hacking: how cybercriminals are using AI coding agents to scale data extortion operations"**
 - Summary:** "Today we are sharing insights about a sophisticated cybercriminal operation (tracked as GTG-2002) we recently disrupted that represents a new evolution in how cyber threat actors leverage AI—using coding agents to actively execute operations on victim networks, known as “vibe hacking”." It details a cybercriminal using Claude Code for a scaled data extortion operation across multiple international targets in a short timeframe, automating reconnaissance, credential harvesting, and network penetration at scale, affecting at least 17 distinct organizations in just the last month across government, healthcare, emergency services, and religious institutions.
 - ABOUT CLAUDE CODE:** "Anthropic's agentic coding tool that lives in your terminal, understands your codebase, and helps you code faster through natural language commands."
 - Key findings:** "Our investigation revealed that the cybercriminal operated across multiple sectors, creating a systematic attack campaign that focused on comprehensive data theft and extortion. The operation leveraged opportunistic targeting based on results from using open source intelligence tools and scanning of internet-facing devices. The actor demonstrated unprecedented integration of artificial intelligence throughout their attack lifecycle, with Claude Code supporting reconnaissance, exploitation, lateral movement, and data exfiltration. The actor provided Claude Code with their preferred operational TTPs (Tactics, Techniques, and Procedures) in their GLAIDE.mdl file that is used as a guide for Claude Code to respond to prompts in a manner preferred by the user. However, this was simply a preferential guide and the operation still utilized Claude Code to make both tactical and strategic decisions—determining how best to penetrate networks, which data to exfiltrate, and how to craft psychologically targeted extortion demands. The actor's systematic approach resulted in the compromise of personal records, including healthcare data, financial information, government credentials, and other sensitive information, with direct ransom demands occasionally exceeding \$500,000."
- THREAT ASSESSMENT Document:** Partially visible at the bottom right.

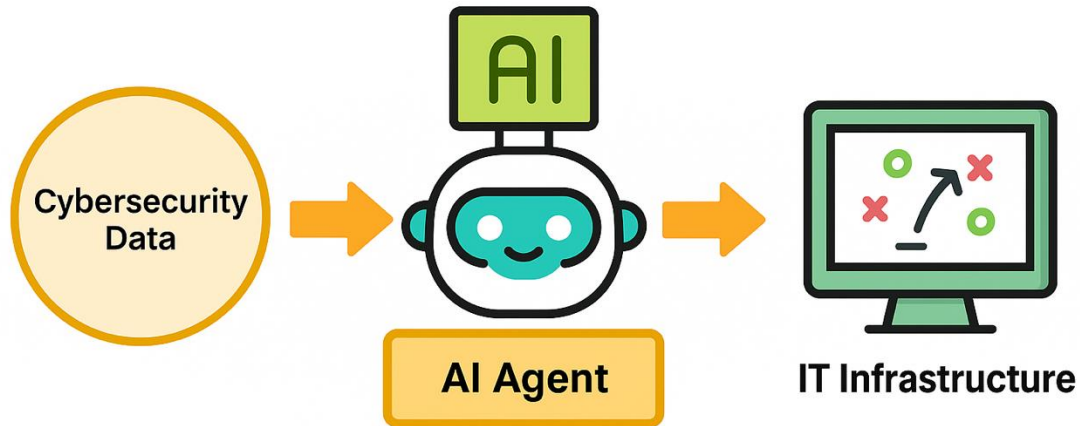
Research Activity (1), Contribution

We focus on the automated generation of complete **PowerShell**-based attacks. The main contributions of this work are:

- Collection of malware-related data
- Construction of a comprehensive dataset
- Automated generation of PowerShell malware
- Systematic assessment of the generated samples



Research Activity (2): Problem statement



Challenge:

- Shared cyber threat intelligence (CTI) platforms often remain underutilized, as they primarily serve as information repositories rather than actionable resources for cyber incident response (IR) .

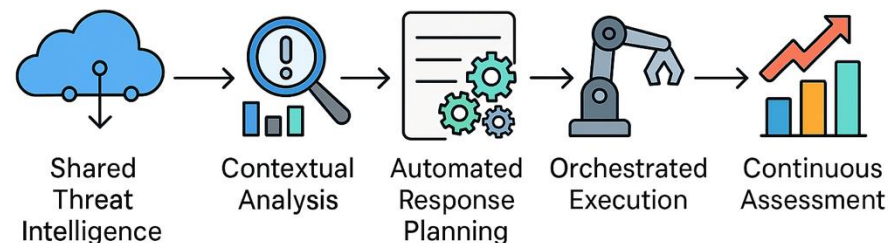
Research Focus:

- We aim to make shared threat intelligence platforms truly actionable by leveraging the Agentic AI paradigm to autonomously orchestrate cyber incident response operations.

Research Activity (2): Contributions

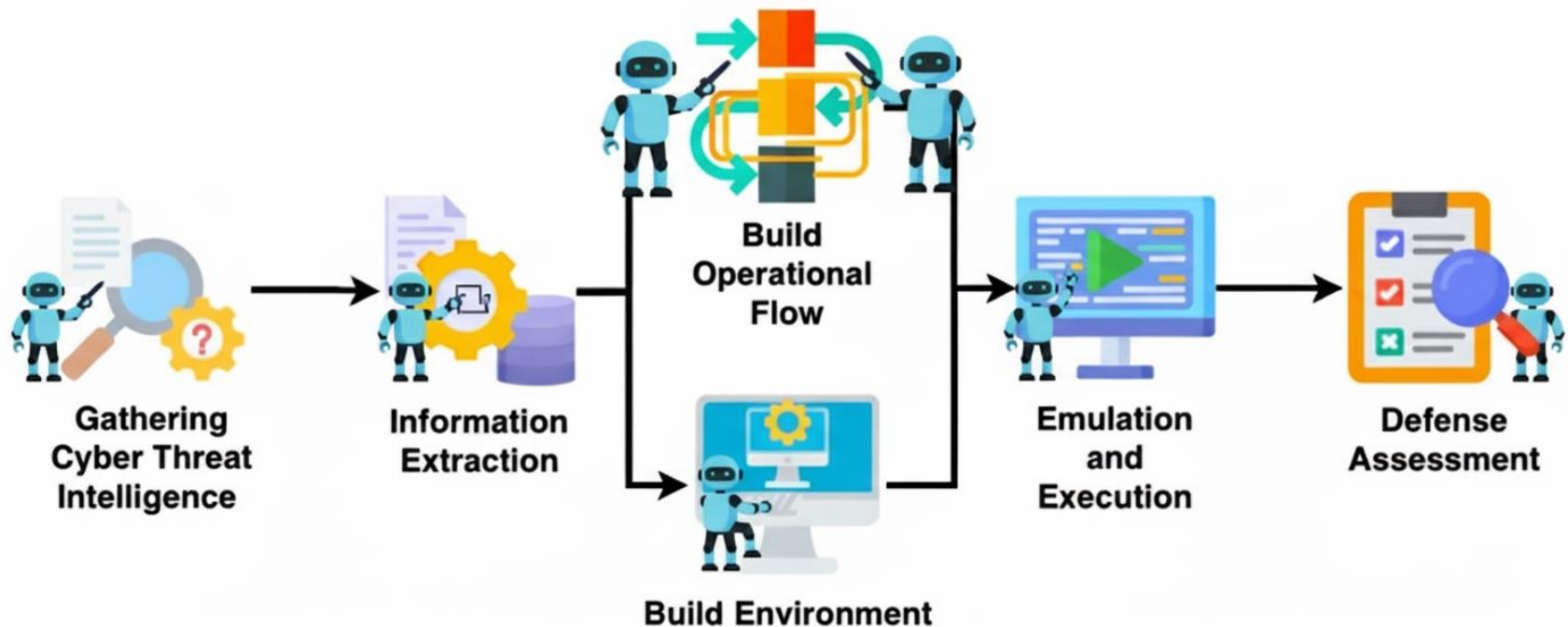
We are actively developing methodologies to operationalize shared threat intelligence for automated incident response.

- Composing **response plans** directly from shared threat intelligence platforms
- Designing and refining **Agentic AI-driven workflows**
- Define **standards** for IR and CTI to enable their use within an Agentic AI paradigm
- Systematically **evaluating** and improving generated response plan



Future Work

Future Work will prioritize the development and evaluation of offensive security methodologies enabled by the Agentic AI paradigm, deepening experimental research and disseminating new findings.



Research products

[C2]	Della Penna, S., Natella, R., Orbinato, V., Parracino, L., & Pianese, L. (2025). CTI-HAL: A Human-Annotated Dataset for Cyber Threat Intelligence Analysis. - Workshop on Attackers and CybeCrime Operations (WACCO) Published
------	--

Thank you for your Attention !